DOCUMENT RESUME

ED 342 797 TM 017 962

AUTHOR Marsh, Herbert W.; Grayson, David
TITLE Multitrait-Multimethod Data: An Evaluation of Five Analytic Approaches.
PUB DATE 15 Oct 91
NOTE 49p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Variance; Comparative Analysis; *Construct Validity; Definitions; Foreign Countries; *Models; *Multitrait Multimethod Techniques; Sample Size
IDENTIFIERS *Campbell Fiske Matrix; *Confirmatory Factor Analysis

ABSTRACT

The multitrait-multimethod (MTMM) design of D. T. Campbell and D. W. Fiske (1959) is the most widely used paradigm for testing construct validity, but it continues to be plagued by problems associated with definition of terms, operationalizations of their guidelines, and analytic procedures used to test them. Using five diverse MTMM data sets, five current analytic approaches are demonstrated, compared, and contrasted. These include two manifest variables approaches: the original guidelines of Campbell and Fiske and the analysis of variance model. Also studied are the following latent variables approaches: (1) a taxonomy of confirmatory factor analysis (CFA) models; (2) the covariance component analysis (CCA) model; and (3) the composite direct product (CDP) model. Even though the five approaches use a common terminology (convergent validity, discriminant validity, and method effects), there is a fuzziness about what the concepts mean and how they are operationalized. This review and analysis recommends a common terminology and operationalization of terms based on CFA models. The use of four CFA models and the CDP model is also recommended, along with the original Campbell-Fiske guidelines. The strongest single model appears to be the CFA correlated uniqueness model. One figure and two tables present study findings, and a 52-item list of references is included. Two appendices provide an additional 10 pages of tables. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Multitrait-multimethod Data: An Evaluation of Five Analytic Approaches

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Herbert W. Marsh
University of Sydney Western, Macarthur

David Grayson
University of Sydney

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
HERBERT W. MARSH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

15 October, 1991

Running Head: Multitrait-multimethod Data

ED342797

ABSTRACT

Campbell and Fiske's (1959) multitrait-multimethod (MTMM) design is the most widely used paradigm for testing construct validity, but it continues to be plagued by problems associated with definitions of terms, operationalizations of their guidelines, and analytic procedures used to test them. Using five diverse MTMM data sets, we demonstrate, compare and contrast, and evaluate five current analytic approaches: two manifest variable approaches (Campbell and Fiske's (1959) original guidelines and the ANOVA model) and three latent variables approaches (a taxonomy of confirmatory factor analysis (CFA) models, the covariance component analysis (CCA) model, and the composite direct product (CDP) model). Even though the five approaches use a common terminology -- convergent validity, discriminant validity, and method effects -- there is a "fuzziness" about what these concepts mean and how they are operationalized in the different approaches. Based on our review and analysis we recommend a common terminology and operationalization of terms based on CFA models, and recommend the use of four CFA models and the CDP model along with the original Campbell-Fiske guidelines. The strongest single model, however, appears to be the CFA correlated uniqueness model.

TM017962

BEST COPY AVAILABLE

Campbell and Fiske (1959) advocated the assessment of construct validity by measuring multiple traits (T1, T2, ..., Tt) with multiple methods (M1, M2, ..., Mm). Traits refer to attributes such as abilities, attitudes, and personality characteristics. In many applications, the multiple traits represent a multidimensional construct (e.g., self-concept) in which there are likely to be moderate to large correlations among the different traits and it may be reasonable to predict a priori the pattern of relations among the different constructs. In some applications, however, the multiple traits are conceptually unrelated (e.g., attitudes towards smoking and capital punishment) so that it may be difficult to predict the pattern of correlations among traits.. The term multiple methods was used very broadly by Campbell and Fiske (also see Fiske, 1982) to refer to multiple tests, multiple methods of assessment, multiple raters, or multiple occasions. The MTMM design is frequently used to study multiple battery data in which the same measures are presented on multiple occasions to study stability, or across different raters to study rater agreement (Browne, 1984; Cudeck, 1988; Marsh, 1989; Wothke & Browne, 1990). Whereas the analytic procedures for evaluating MTMM data are appropriate for different types of multiple measures, the substantive interpretations differ depending on the nature of the multiple methods. It is also evident that the extent of support for the construct validity of responses associated with any particular trait or method will depend in part on the other traits and methods that are included in the design.

In evaluating multitrait-multimethod (MTMM) data it is typical to refer to convergent validity, discriminant validity, and method effects (Campbell & Fiske, 1959; Marsh, 1988). Convergent validity refers to true score or common factor trait variance. In the Campbell-Fiske approach it is inferred from agreement between measures of the same trait assessed by different methods -- the convergent validities. Discriminant validity refers to the distinctiveness of the different traits. In the Campbell-Fiske approach it is inferred by comparing correlations among different traits to the reliabilities of the traits and to convergent validities. Method effects refers to the influence of a particular method and is typically viewed as an undesirable bias that inflates the correlations among the different traits that are measured by the same method (but also see Campbell & O'Connell, 1982). In the Campbell-Fiske approach it is inferred by comparing correlations among traits measured by the same method with correlations among the same traits measured by different methods.

The Campbell-Fiske MTMM paradigm is, perhaps, the most widely employed construct validation design, and their original guidelines remain the most frequently used approach for examining MTMM data. However, important problems with their guidelines are well known (e.g., Althauser & Heberlein, 1970; Alwin, 1974; Campbell & O'Connell, 1967; Marsh, 1988; 1989; Wothke, 1984, 1987) and have led to many alternative analytic approaches. Kenny and Kashy (in press) noted that even after 30 years of widespead use, we still do not know how to analyze adequately data resulting from the MTMM paradigm. Early attention was received by an ANOVA model proposed by Stanley (1961; also see Kavanagh, Mackinney & Wollins, 1971; Marsh, 1988; Marsh & Hocevar, 1983). Subsequently, considerable attention was given to confirmatory factor analysis (CFA) approaches (Joreskog, 1974; Widaman, 1985; Marsh, 1988; 1989). However, researchers have pointed to what appears to be an inherent instability in the general CFA model due, perhaps, to empirical underidentification, such that this model usually results in improper solutions. Partly in response to this problem, researchers have demonstrated the use of different approaches that are more likely to result in proper solutions: a different CFA model called the correlated uniqueness model (Marsh, 1988; 1989; also see Kenny, 1976; Kenny & Kashy, in press; also see Browne, 1980); a covariance component analysis (CCA; Wothke, 1984, 1987; also see Browne, 1989; Kenny & Kashy, in press); and the composite direct product model (CDP; Browne, 1984; 1989; also see Cudeck, 1988; Bagozzi & Yi, 1990; Wothke &

Browne, 1990). Whereas each of the approaches has potential strengths weaknesses, there has been insufficient attention given to comparing the different analytic strategies.

The purpose of this study is to demonstrate, compare and contrast, and evaluate five approaches to evaluating MTMM data. Two of the approaches are based on relations among manifest variables; Campbell and Fiske's (1959) original guidelines and an ANOVA model proposed by Stanley (1961). The remaining three approaches are based on relations among latent variables -- a taxonomy of confirmatory factor analysis (CFA) models (Widaman, 1985; Marsh, 1989), Wothke's (1984, 1987) covariance component analysis (CCA) model, and Browne's (1984, 1989) composite direct product (CDP) model. It must be emphasized that the five approaches are not equivalent in their operationalizations of the terms convergent validity, discriminant validity, and method effects. Consequently, different approaches will sometimes result in incompatible interpretations. Also, each approach has different strengths and weaknesses that may be idiosyncratic to particular applications. In this respect, it is important to evaluate the different approaches using a wide variety of MTMM studies.

Surprisingly, no previously published research has compared results from the five approaches considered here, or even the three latent variable approaches that are our primary emphasis. Furthermore, the correlated uniqueness model, which we argue is the strongest model in the CFA taxonomy, has not been systematically compared with either the CCA or the CDP approach. To remedy this situation, we apply all five approaches to a set of five MTMM studies specifically selected to represent a variety of different MTMM designs and outcomes. This breadth of application is important because most previous research has compared one or, in a very few cases, two, of the latent variable approaches with the traditional Campbell-Fiske guidelines for a single set of data. Ours is apparently the first to apply such wide variety of approaches to such a diversity of MTMM studies. After briefly describing the five MTMM matrices, we apply the manifest variable approaches, describe the three latent variable approaches, and compare results based on the latent variable approaches.

A Description of Five MTMM Matrices

For purposes of the present investigation, we have chosen 5 MTMM matrices (see Appendix 1) that represent a variety of different MTMM designs and patterns of results. Because these matrices are based on previously published data in which the methodological details are presented in greater detail, we offer only brief summaries here.

Byrne Data. Byrne and Shavelson (1986; also see Marsh, 1988; 1989; Marsh, Byrne & Shavelson, 1988) examined the relations between three academic self-concept traits (Math, Verbal, and General School) measured by three different instruments. The 9 scores representing all combinations of the 3 traits and 3 methods were based on multi-item scales and the three instruments had strong psychometric properties. Consistent with theory and considerable prior research, it was found that the Math and Verbal self-concepts were nearly uncorrelated with each other and were substantially correlated with School self-concept. Marsh (1989) noted that this "is an exemplary MTMM study because of the clear support for the Campbell-Fiske guidelines, the large sample size (817, after deleting persons with missing data), the good psychometric properties of the measures, and the a priori knowledge of the trait factor structure" (p. 348). Also, the predicted lack of correlation between Math and Verbal self-concept satisfies the Campbell and Fiske recommendation to include two traits "which are postulated to be independent of each other" (p. 104). In the 3Tx3M design, apparently comparable traits were inferred from responses to different instruments completed by the same individuals.

Kelly and Fiske Data. This MTMM matrix is one of those originally considered by Campbell and Fiske (1959). Kelly and Fiske (1951) examined relations among ratings of 124 first-year clinical psychology students by

the students themselves, by the median response from a set of three other students, and by the pooled ratings of the assessment staff. The multiple traits were "behaviors that can be observed directly on the surface" (Campbell & Fiske, 1959). For purposes of illustration, Campbell and Fiske selected five traits that best represented underlying factors found in separate analyses of ratings by each group (assertive, cheerful, serious, unshakable poise, and broad interests). Campbell and Fiske (1959), noting a lack of support for their guidelines in most MTMM studies, concluded that this MTMM matrix "is, we believe, typical of the best validity in personality trait ratings that psychology has to offer at the present time." This matrix was also the basis of Joreskog's original (1974) presentation of the CFA approach, although the general model that he proposed actually resulted in an improper solution. Browne (1984) demonstrated his CDP model with this matrix and claimed that it was superior to the CFA approach. Wothke (1984, 1987) demonstrated his CDP model with this matrix, also claiming its superiority over the CFA approach. However, the CFA correlated uniqueness model was not considered in either of these studies, so the claimed superiority of either the CCA or CDP approach over the CFA approach for this data may be premature. In the 5Tx3M design, apparently comparable traits were inferred from responses to different stimulus materials by different individuals who have different roles.

Ireland Data. Marsh and Ireland (1984, 1988) asked multiple teachers to evaluate 139 student essays according to different components of writing effectiveness. In this application, responses by three teachers constitute the multiple methods of assessing 5 traits (mechanics, sentence structure, word usage, organization, content/ideas, quality of style). Marsh and Ireland found good support for agreement among different teachers (and between teacher ratings and school based measures) for the total scores. Consistent with previous research, however, they found little or no support for the ability of teachers to differentiate among the multiple traits. This MTMM matrix, then, represents a multiple battery design (the same stimulus material was used for multiple raters) in which there is apparently good support for convergent validity, but not discriminant validity. Marsh (1989) also demonstrated the CFA correlated uniqueness model with this data, arguing for its superiority to the general CFA model that resulted in an improper solution. In the 5Tx3M design, apparently comparable traits were inferred from responses to the same stimulus materials by different individuals who have the same role.

Youth In Transition (YIT) Data. Data for this matrix come from the YIT study (Bachman, 1975) in which a large, nationally representative sample of high school males were sampled in 10th, 11th, and 12th grades and one year after graduation from high school. For this data, the multiple occasions are considered to be the different methods of assessment. Marsh and Bailey (1991) found strong support for convergent and discriminant validity, and weak method (occasion) effects for a large number of variables from this data. For present purposes we consider five traits (self-esteem, political knowledge, honesty, job ambition, and anxiety) measured on three different occasions. This MTMM matrix, then, represents multiple battery data (same measures on different occasions) in which there is good support for both convergent and discriminant validity. In the 5Tx3M design, apparently comparable traits were inferred from responses to the same stimulus materials by the same individuals on different occasions.

Lawler data. Lawler (1969) considered ratings of three job performance traits (quality of job performance, ability to perform job, and effort put into the job). The multiple methods were self-ratings, ratings by supervisors, and ratings by peers. This matrix, along with the one based on the Kelly and Fiske data, has apparently been the most frequently reanalyzed MTMM matrix, but the results have not been clear-cut. There is moderate agreement between peers and supervisors, but also substantial correlations among the different traits. Self-ratings of effort are moderately correlated with those of peers and supervisors, but self-ratings on the other two traits are nearly uncorrelated with the peer and supervisor ratings. This 3Tx3M matrix, then represents typically "messy" data. In the 3Tx3M design,

apparently comparable traits were inferred from responses to the same stimulus materials by different individuals who have different roles.

In summary, we selected MTMM matrices that vary substantially in design and in terms of apparent support for convergent validity, discriminant vaiidity, and method effects.

The Campbell and Fiske (1959) Approach

The MTMM matrix based on the Byrne data (Appendix 1) is used to illustrate the MTMM terminology that is embodied in the Campbell-Fiske approach. There are three traits, School self-concept (T1), Verbal self-concept (T2) and Math self-concept (T3), and three methods -- three different self-concept instruments (M1, M2, M3). The MTMM matrix contains correlations among these M x T = 9 measures. The measured variables are typically ordered in terms of traits within method (e.g., T1M1, T2M1, T3M1, T1M2, ... T3M3). The MTMM matrix is divided into triangular submatrices of relations among measures assessed with the same method (monomethod), and square submatrices of relations among measures assessed with different methods (heteromethod). Adopting the Campbell and Fiske terminology, there are four types of coefficients: (a) monotrait-monomethod coefficients or reliability estimates, the values in parentheses along the main diagonal of the MTMM matrix or 1.0s if no reliability estimates are available; (b) heterotrait-monomethod (HTMM; different traits, same method) coefficients, the off-diagonal coefficients of the triangular submatrices; (c) monotrait-heteromethod (MTHM; same traits, different method) coefficients or convergent validities, the values in the diagonals of the square submatrices; and (d) heterotrait-heteromethod (HTHM; different traits, different methods) coefficients, the off-diagonal coefficients of the square submatrices.

Campbell and Fiske (1959) proposed four guidelines for evaluating MTMM matrices and inferring support or nonsupport for convergent and discriminant validity, although they actually suggested other possible guidelines. The application of the guidelines is presented in detail for the Byrne data, whereas the application to the other MTMM matrices (see Appendix 1) is summarized in Table 1.

For purposes of explanation, manifest score are denoted x(Ti, Mp) where Ti is one of the multiple traits (Ti, Tj, Tk...) and Mp is one of the multiple methods (Mp, Mq, Mr...). Let r(TiMp, TjMs) denote the correlation between x(Ti, Mp) and x(Tj, Ms).

Convergent validity criterion

1) r (TiMp, TiMs) >> 0

The convergent validity coefficients should be statistically significant and sufficiently large to warrant further examination of validity. Failure of this criterion suggests that different measures are measuring different constructs, implying a lack of validity for at least some of the measures, or that true trait variance is small relative to the size of method effects and measurement error. Although positive convergent validity coefficients may also reflect shared method effects, satisfaction of this guideline is a logical prerequisite to the consideration of other guidelines. For the Byrne data all nine convergent validities are statistically significant, varying between .54 and .87 (mean r = .70), thus providing strong support for this guideline.

Discriminant validity guidelines

2) r (TiMp, TiMq) > r (TjMp, TiMq) and r (TiMp, TiMq) > r (TjMq, TiMp), p not equal p

The convergent validities should be higher than HTHM correlations. The failure of this criterion implies that agreement on a particular trait is not independent of agreement on other traits, suggesting that agreement can be explained by true trait correlations or shared method effects. For T=3 and M=3 this criterion requires each convergent validity to be higher than the 4 HTHM coefficients in the same row and column of the square submatrix. Because

convergent validities (mean r = .70) are higher than the comparison correlations (mean r = .31) in all 36 of these comparisons, there is good support for this guideline of discriminant validity for the Byrne data.

3) r (TiMp, TiMq) > r (TiMp, TjMp) and r (TiMp, TiMq) > r (TiMq, TjMq)

The convergent validities should be higher than HTMM correlations. Violations of this criterion suggest that there are true trait correlations and/or method effects. Particularly if HTMM correlations approach the reliability estimates then there is evidence that the traits are not measuring different constructs and/or a strong method effect. This criterion requires each convergent validity to be higher than the 4 HTMM comparison coefficients in the same row and column of the corresponding triangular submatrices. Because the convergent validities (mean r = .70) are higher than the comparison correlations (mean r = .35) for 33 of 36 comparisons, there is reasonable support for this criterion in the Byrne data. All three failures involve M3 where correlations among the traits (mean r = .44) are higher than for M1 (.28) or M2 (.33).

4) r (TiMp, TjMq) > r (TkMp, TlMq) implies r (TiMr, TjMs) > r (TkMr, TlMs).

The pattern of correlations among traits should be similar for the same and different methods. Assuming that there are significant correlations, satisfaction of this criterion suggests true trait correlations that are independent of the method of assessment whereas failure suggests that the observed correlations are differentially affected by method effects. When the number of traits is small this criterion is typically examined by inspection of the rank order of correlations (e.g., Sullivan and Feldman, 1979), but Marsh (1982) correlated the correlations to obtain a more precise index of similarity when the number of traits was large. The relative size of correlations within each method correlated between .66 and .67 with the corresponding correlations within the other methods. All correlations between Math and Verbal self-concepts are small (mean r = .06) whereas school self-concept is significantly and consistently correlated with both Math (mean r = .45) and Verbal (mean r = .42) self-concepts. These results, particularly since they support a priori hypotheses about the pattern of correlations, provide clear support for this guideline in the Byrne data.

5) $r (TiMp, TjMp)/ [(r (TiMp, TiMp) r(TjMp, TjMp)]^{1/2} << 1$

Campbell and Fiske (1959) specifically stated that a clear violation of discriminant validity occurred "where within a monomethod block, the heterotrait values are as high as the reliabilities" (p. 84) and that "the elevation of the reliabilities above the heterotrait-monomethod triangle is further evidence for discriminant validity" (p. 97). Although not formally included as one of their guidelines, it is clear that this was part of their strategy for evaluating MTMM matrices. In retrospect, its exclusion from their "official" list of guidelines is unfortunate, because it would have encouraged researchers to systematically evaluate the reliability of their measures, to focus more on the quality of measurement of each trait-method unit, to evaluate the implicit assumption of equally reliable measures underlying all the guidelines, and to include this as part of the MTMM matrix. We realize that this guideline cannot be evaluated in most existing MTMM studies because reliability estimates typically are not reported, but have presented it as one of the Campbell-Fiske guidelines to encourage its consideration in future research. For the Byrne data the coefficient alpha estimates of reliability (.79 to .95; mean = .89) are all substantial, and none of the disattenuated correlations approaches 1.0, providing good support for this guideline.

Method effects

6) r (TiMp, TjMp) > r (TiMp, TjMq)

Campbell and Fiske (1959, p. 85) stated that "the presence of method variance is indicated by the difference in level of correlation between parallel values of the monomethod block and the heteromethod block, assuming comparable reliabilities among the tests." Large differences imply substantial method effects and/or shared method effects.

Although not formally included as one of their guidelines, Marsh (1988; also see Millsap, 1990) noted that this was an important aspect of their approach and proposed its addition to the set of guidelines. In operationalizing this criterion for the Byrne data, for example, the mean HTMM correlation is .35 whereas the mean HTHM correlation is .29, suggesting a small amount of method effect. The correlations among traits, however, are larger for M3 (mean r = .44), than for M2 (mean r = .33) and M1 (mean r = .28). This suggests modest amounts of method effect for M3, but little or no method effects for M1 and M2.

Insert Table 1 About Here

The application of the Campbell-Fiske guidelines (except for guideline 5 that requires reliability estimates that are typically unavailable) for the other four data sets is summarized in Table 1. The convergent validities are consistently large in the YIT (mean r = .52) and Ireland (mean r = .62) data, but less so in the Kelly and Fiske (mean r = .36) and Lawler (mean r = .28) data sets. Guidelines 2 and 3 are satisfied for most comparisons for the YIT and for the Kelly and Fiske data, but not for the Lawler or the Ireland data. The patterns of correlations among traits is reasonably similar across methods in all 4 data sets. Method effects, based on the comparison of HTMM and HTHM correlations, appear to be substantial for the Lawler and Ireland data, but not for the Kelly and Fiske and the YIT data. In summary, the Campbell-Fiske guidelines appear to be well satisfied for the Byrne, YIT, and, to a lesser extent, for the Kelly-Fiske data. There is good support for only convergent validity for the Ireland data and even the support for convergent validity is weak for the Lawler data.

Problems With the Campbell Fiske Guidelines

The Campbell-Fiske guidelines continue to be widely used and are useful in many instances. Because of their popularity, ease of application, intuitive appeal, heuristic value, and wide recognition, it is recommended that these guidelines should be applied as an initial step in MTMM studies even though more sophisticated approaches should also be used. If inferences based on the Campbell-Fiske guidelines do not agree with those based on other analytic approaches, then the appropriateness of both approaches should be more fully examined. This requires researchers to better understand the different approaches. The following issues represent important limitations to the Campbell-Fiske guidelines, some of which are addressed by other approaches.

The number of comparisons. For the 3Tx3M design, guidelines 2 and 3 required a total of 72 comparisons between convergent validities and other correlations. However, these comparisons are not tests of statistical significance and appropriate significance tests would be difficult to devise for so many nonindependent comparisons. Furthermore, the number of comparisons goes up geometrically with the number of traits and methods. For example, 3164 comparisons are required for a 12Tx4M design (Marsh, Barnes & Hocevar, 1985). The researcher must then decide whether the proportion of failures is sufficiently low, whether mean difference between convergent validities and comparison coefficients is sufficiently large, or whether size and pattern of violations are sufficiently unsystematic to warrant support of a criterion. This decision is somewhat arbitrary.

Correlated traits and discriminant validity. Support for discriminant validity should, apparently, be based on the size of true trait correlations. If, for example, true trait correlations approach 1.0 or exceed some arbitrary value, then the traits could be said to lack of discriminate validity. Campbell and Fiske distinguish between method variance, true trait variance, and true trait covariance. Method variance associated with a particular method of assessment is detrimental to discriminant validity in the Campbell-Fiske guidelines, but does not preclude it. True trait variance, inferred from the correlation between different measures of the same trait that is independent of method variance, is good but does not imply discriminant validity. True trait covariation, the true correlation between

different traits that is independent of method effects, will increase the likelihood of failures of guidelines 2 and 3. However, criterion 4 specifically tests for true trait covariation and is interpreted as support for discriminant validity. A complete lack of true trait covariation or trait correlations approaching the reliability of the measures makes interpretation simple, but is unlikely. Hence, true trait correlations and their interpretation in relation to discriminant validity is ambiguous within the Campbell-Fiske approach.

Inferences based on observed correlations and errorful data. The validity of inferences based on the Campbell-Fiske guidelines depends on the behavior of the underlying constructs, but the Campbell-Fiske guidelines are applied to correlations between observed measures. Campbell and Fiske noted that the application of their guidelines implicitly assumes the measures to be equally reliable. If the reliabilities differ substantially, then inferences based on the guidelines may be invalid. For example, correlations among traits assessed with a more reliable method may produce higher trait correlations than a less reliable method, and thus give the impression of larger method effects. Other researchers have attempted to evaluate what assumptions about underlying constructs are required in order for inferences based on the guidelines to be valid (e.g., Althauser & Heberlein, 1970; Alwin, 1974; Marsh, 1988; Sullivan and Feldman, 1979). There is, however, neither clear agreement about what conditions invalidate the inferences nor practical solutions about how to evaluate these inferences.

Large method effects and shared method effects. Whereas the Campbell-Fiske guidelines were designed to test for convergent and discriminant validity when method effects are likely, the existence of large method effects and shared method effects may undermine interpretations of the guidelines. Thus, for example, large method effects will lead to what appears to be a lack of discriminant validity (according to guidelines 2 and 3) even when the underlying traits are distinct. High convergent validities may also reflect substantial shared method effects in addition to, or instead of, true trait effects that generalize across methods. If different method effects are negatively correlated, a zero convergent validity could reflect the counter-balancing negative shared method effects and positive true trait variance. Even the fifth criterion used to infer the size of method effects must be interpreted cautiously when there are large shared method effects. In the extreme, if all the method effects are large and correlations between method effects representing different methods approach 1.0, then application of the fifth criterion would imply a lack of method effects. In this sense, inferences based on the Campbell-Fiske guidelines should be interpreted as evidence about the trait effects relative to the size of method effects. Whereas large method effects and shared method effects make it difficult to make inferences about true trait variance and true trait covariance in the Campbell-Fiske approach, this may not be a crippling problem. From a practical perspective, if the method effects are huge, then the validity of the interpretations of the relatively tiny trait effects may not be very important.

Trait/method correlations and interactions. Interpretations of the discriminant validity guidelines summarized above are based on the assumption that traits are uncorrelated with method effects. While this assumption may be substantively reasonable in some applications, its justification is primarily pragmatic rather than substantive. Without such an assumption the interpretation of the guidelines is more complicated and apparently more problematic, but the effect of its violation on the inferences is not well documented (see Althauser & Heberlein, 1970; Wothke, 1984). Campbell and O'Connell (1967) also proposed that traits and methods may interact. Trait/method interactions are different from trait/method correlations. Trait/method correlations imply that there is an overlap in the variance that can be explained by the main effects of traits and methods, whereas trait/method interactions imply that additional variance can be explained by trait/method crossproducts. Whereas the existence of trait-method correlations further complicate the interpretation of the Campbell-Fiske guidelines, the existence of trait-method interactions apparently

undermines the logical basis for the guidelines, the assumption of additivity underlying factor analysis in general, and, perhaps, even the classical approach to test theory (Campbell & O'Connell, 1967; 1982).

In Defense of Campbell and Fiske's Intentions

The subsequently popularized factor analysis representation of MTMM data was apparently the basis of the guidelines proposed by Campbell and Fiske (1959) and subsequently described in Campbell and O'Connell (1967, 1982; also see Kenny & Kashy, in press). Campbell and Fiske specifically noted that "each test or task employed for measurement purposes is a trait-method unit, a union of a particular trait content with measurement procedures not specific to that content. The systematic variance among test scores can be due to responses to the measurement features as well as responses to the trait content" (p. 81). Elsewhere they endorsed Cronbach's (1946, p.475) statement that "the final score ... is a composite of effects resulting from the content of the item and effects resulting from the form of the test used." Campbell and O'Connell (1967) subsequently considered hypothetical MTMM results constructed by varying aspects of latent trait factor loadings, latent method factor loadings, uniqueness, and the associated variance components. Kenny and Kashy (in press) are even more emphatic in making this point, stating that "this [general CFA] model is particularly attractive in that its structure directly corresponds to Campbell and Fiske's original conceptualization of the MTMM matrix" (p. 5).

From this perspective, it is important to emphasize that Campbell and Fiske (1959) explicitly or implicitly noted most of the problems that have been raised in relation to a strict interpretation of their guidelines. Their guidelines, however, were apparently not intended to be given such a strict interpretation nor to be the rigid, inflexible criteria that they have come to represent. Instead, Campbell and Fiske viewed the guidelines as "common-sense desideratum" (p. 83) and suggested that formal statistical analyses such as factor analyses "are neither necessary nor appropriate at this time" (p. 103). They argued that "we believe that a careful examination of a multitrait-multimethod matrix will indicate to the experimenter what his next steps should be: it will indicate which methods should be discarded or replaced, which concepts need sharper delineation, and which concepts are poorly measured because of excessive or confounding method variance" (p. 103). More recently, Fiske (1982) reiterated this contention, adding that "I continue to believe that direct inspection of each trait-method unit should be carried out in every instance. With a little thought and practice, the major interpretations of the matrix will become apparent to the investigator" (p. 80). Their intent apparently was to provide a systematic approach to the formative evaluation of MTMM data at the level of the individual trait-method unit, qualified by the recognized limitations of their approach, not to provide abstract, global summaries of convergent validity, discriminant validity and method effects that are a definitive summative statement. We argue strongly that this formative orientation in the MTMM paradigm must not be lost in the development of mathematically more sophisticated approaches to MTMM data -- that the baby should not be thrown out with the bath water -- and propose that alternative approaches should be evaluated in relation to this original orientation. More generally, Campbell and Fiske had a heuristic intention to encourage researchers to consider the concepts of convergent validity, discriminant validity, and method effects; in this intention the were unquestionably successful.

In summary, the Campbell-Fiske approach provides a heuristic, potentially useful structure for the formative evaluation of MTMM data. However, as acknowledged by Campbell and Fiske (1959), there are many potentially serious problems and ambiguities in the interpretation of their guidelines. The heuristic importance of their work as well as limitations in their guidelines have led to the development of alternative approaches to the evaluation of MTMM data that are considered here.

The ANOVA Model for MTMM Data

MTMM data can be analyzed with a three-factor unreplicated ANOVA and the ANOVA terms can be computed directly from the MTMM matrix (Kavanagh, MacKinney & Wollins, 1971; Marsh & Hocevar, 1983; Schmitt & Stults, 1986). When measures for all levels of traits and methods are obtained for the same subject, three orthogonal sources of variation can be estimated. The main effect of subjects is a test of whether there are significant differences between subjects for measures averaged across traits and methods, and is used to infer convergent validity. The subject x trait interaction tests whether differences between subjects depend on traits, and is used to infer discriminant validity. If it is nonsignificant then the traits have no differential validity in that subjects are ranked the same for all traits. The subject x method interaction tests whether differentiation depends on the method of assessment, and is used to infer method effects. If it is significant then the method effects introduce a systematic source of what is usually interpreted to be an undesirable variance. The three-way interaction is assumed to reflect only random error such that differentiation does not depend on specific trait-method combinations. The main effects due to traits and methods are rarely of substantive interest and are necessarily zero for standardized data. Whereas there are nominal tests of statistical significance for the effects used to infer convergent validity, discriminant validity, and method effects, the primary interest is typically in variance components associated with these effects.

The computation of effects and variance components is described by Kavanagh, MacKinney and Wollins (1971) and by Marsh and Hocevar (1983), and results for the 5 MTMM matrices are presented in Table 1. According to this approach, the variance components associated with convergent and discriminant validity are both substantially larger than the variance component associated with method effects for the Byrne, the YIT, and the Kelly-Fiske data. For the Lawler data the effects of convergent validity and method effects are large, but the discriminant validity effect is small. For the Ireland matrix, the convergent validity effect is very large, the method effect is small, and the discriminant validity effect is very small.

Problems With the ANOVA approach

The advantages of the ANOVA approach are its ease of application and the convenient summary statistics used to infer convergent, discriminant, and method/halo effects. The ANOVA model provides only a global evaluation of variance components and fails to provide the formative evaluation of specific trait-method units that was an original intent of the MTMM paradigm.

The effects in ANOVA model bear some resemblance to terms used in the Campbell-Fiske approach, but it is important to emphasize that they are not directly comparable. In the ANOVA approach, for example, convergence is based on the average correlation in the entire MTMM matrix, whereas in the Campbell-Fiske approach it is based on just the convergent validities. Thus, for example, the Ireland matrix has a much larger convergent validity effect than any of the matrices according to the ANOVA approach even though the mean convergent validity is highest in the Byrne matrix. Also, the Lawler matrix has the third highest (of 5) convergent validity effect in the ANOVA approach, but has the lowest mean convergent validity. In the ANOVA approach, an extremely high convergent validity effect precludes strong support for discriminant validity, whereas strong convergent validity is a prerequisite to discriminant validity in the Campbell-Fiske approach. Because of these disjunctures in terminology in the two approaches, interpretations based on the ANOVA approach should be described carefully so as to not confuse them with the more prevalent Campbell-Fiske terminology. It is also worth noting that the ANOVA approach is sensitive to the orientation of the traits. Thus, for example, if all the traits are positively correlated, then reversing the sign of correlations associated with one particular trait will reduce the average correlation among all traits which will reduce the convergent validity effect and increase the discriminant validity effect.

The ANOVA model cannot be recommended. Like the original Campbell-Fiske guidelines, the ANOVA model is based on inferences about measured, errorful data. Important limitations of the ANOVA model may be overlooked in the model's apparent but deceptive simplicity and precision. Also, this approach does not lead to the heuristic interpretations of specific measures, traits, and methods that may be the most important contribution of the MTMM paradigm as a formative tool. The unfortunate linking of the ANOVA effects to the Campbell-Fiske terminology is inappropriate. The convergent, discriminant, and method/halo effects in the ANOVA model are not the same as those inferred from the Campbell-Fiske guidelines even though the two approaches may lead to apparently consistent conclusions (see Marsh & Hocevar, 1983). The interpretation of the average correlation in the entire MTMM as support for convergent validity is, apparently, particularly dubious. The sensitivity to trait orientation also appears to be a potential problem. At least some of the inherent weaknesses in the ANOVA model are overcome in the related CCA model developed by Wothke (1984, 1987) that is described latter. In this sense, the ANOVA model may have been superseded by Wothke's work.

The Confirmatory Factor Analysis (CFA) Approach

MTMM matrices, like other correlation matrices, can be factor analyzed to infer the underlying dimensions. Factors defined by different measures of the same trait suggest trait effects, whereas factors defined by measures assessed with the same method suggest method effects. With CFA the researcher can define models that posit a priori trait and method factors, and test the ability of such models to fit the data. However, critical problems in the CFA approach are the assumptions underlying the proposed models, technical difficulties in the estimation of parameters, and the validity of inferences based on the parameter estimates (Marsh, 1989).

The CFA approach to MTMM data is the most widely applied alternative to the Campbell-Fiske guidelines. In the general MTMM model adapted from Joreskog (1974; also see Marsh, 1988; 1989; Widaman, 1985): (a) there are at least three traits (T=3) and 3 methods (M=3); (b) T x M measured variables are used to infer T + M a priori factors; (c) each measured variable loads on one trait factor and one method factor but is constrained so as not to load on any other factors; (d) correlations among trait factors and among method factors are freely estimated, but correlations between trait and method factors are fixed to be zero; (e) the uniqueness of each scale is freely estimated but assumed to be uncorrelated with the uniquenesses of other scales. This general model, which we refer to as the CFA model with correlated traits and correlated methods (CFA-CTCM), is presented (Model 1 in Figure 1) for a 4Tx4M design.

Insert Figure 1 About Here

An advantage of this general CFA model is the apparently unambiguous interpretation of convergent validity, discriminant validity, and method effects: large trait factor loadings indicate support for convergent validity, large method factor loadings indicate the existence of method effects, and large trait correlations -- particularly those approaching 1.0 -- indicate a lack of discriminant validity. Also, in standardized form, the squared trait loading, the squared method factor loading, and the error component sum to 1.0 and can be interpreted as components of variance for each item. Again, however, it is important to emphasize that these effects are not the same as the convergent, discriminant, and method effects inferred from the Campbell-Fiske approach. The most obvious difference is that inferences are based on latent constructs instead of manifest variables. Also, as noted earlier, large method effects and correlated method effects can influence interpretations of convergent validity and discriminant validity with the Campbell-Fiske guidelines. Consistent with Kenny and Kashy's (in press) assertion, our interpretation of Campbell and Fiske (1959; also see Campbell & O'Connell, 1967; 1982) suggests that their original guidelines were implicitly based on a latent trait model like the CFA models. From this perspective, the operationalizations of convergent

validity, discriminant validity, and method effects in the CFA approach may better reflect Campbell and Fiske's (1959) original intentions than do their own guidelines.

Researchers have proposed many variations to the CFA-CTCM model to examine inferences about trait or method variance or to test substantive issues specific to a particular study (e.g., Joreskog, 1974; Marsh, Barnes & Hocevar, 1985; Marsh, 1989; Widaman, 1985). Widaman proposed a taxonomy of models that systematically varied different characteristics of the trait and method factors that was expanded by Marsh (1988, 1989). This taxonomy is designed to be appropriate for all MTMM studies, to provide a general framework for making inferences about the effects of trait and method factors, and to objectify the complicated task of formulating models and representing the MTMM data. Whereas detailed consideration of the taxonomy is beyond the scope of the present investigation (see Marsh, 1989), four models (Figure 1) are considered that we recommend as the minimum set of models that should be applied in all CFA MTMM studies.

The trait-only model (CFA-CT; Figure 1) posits trait factors but no method effects whereas the remaining models posit trait factors in combination with different representations of method effects. Hence, the trait-only model is nested under the other CFA models so that the comparison of its fit with the other CFA models provides an indication of the size of methods effects. Implicit in this operationalization of method effects is Joreskog's contention that "method effects are what is left over after all trait factors have been eliminated" (1971, p. 128; also see Marsh, 1989). The model with correlated trait factors but uncorrelated method factors (CFA-CTUM; Figure 1) differs from the CFA-CTCM model only in that correlations among the method factors are constrained to be zero. Hence the comparison of the CFA-CTCM and CFA-CTUM models provides a test of whether method factors are correlated.

In the correlated uniqueness model (CFA-CTCU; Figure 1), method effects are inferred from correlated uniquenesses among measured variables based on the same method instead of method factors (see Marsh, 1989; Marsh and Bailey, 1991; Kenny, 1979; Kenny & Kashy, in press). Like the CFA-CTUM model the CFA-CTCU model assumes that effects associated with one method are uncorrelated with those associated with different methods. The CFA-CTCU models differs from the CFA-CTCM and CFA-CTUM models in that the latter two models implicitly assume that the method effects associated with a given method can be explained by a single latent method factor (hereafter referred to as the unidimensionality of method effects) whereas the correlated uniqueness model does not. This important distinction, however, is only testable when there are at least four traits. When there are three traits the CFA-CTUM and the CFA-CTCU models are equivalent so long as both models result in a proper solution (i.e., the number of estimated parameters goodness of fit are the same fit, and parameter estimates from one can be transformed into the other) because correlations among three indicators can be represented by a single latent trait.

The juxtaposition of the CFA-CTUM, CFA-CTCM, and CFA-CTCU models is important. So long as all three models result in proper solutions, the comparison of CFA-CTUM and CFA-CTCU model tests the unidimensionality of method effects (i.e., whether the method effects associated with each method form a single latent method factor), whereas the comparison of the CFA-CTUM and CFA-CTCM models tests whether effects associated with different methods are correlated. Because the CFA-CTCU and CFA-CTCM are not nested, their comparison is more complicated. For example, if both the CFA-CTCM and CFA-CTCU models fit the data substantially better than the CFA-CTUM, all three models may be wrong: the CFA-CTUM is wrong because it assumes that the effects associated with each method are unidimensional and unrelated to the effects associated with other method; the CFA-CTCM is wrong because it assumes that the effects associated with each method are unidimensional; the CFA-CTCM is wrong because it assumes that the effects associated with each method are unrelated to the effects associated with other methods

From a practical perspective, the most important distinction between the CFA-CTCM, CFA-CTUM, and the CFA-CTCU models is that the CFA-CTCM model typically results in improper solutions, the CFA-CTUM model often results in an improper solution, and the CFA-CTCU almost always results in proper solutions (Kenny & Kashy, in press; Marsh, 1989; Marsh & Bailey, 1991; also see Wothke, 1984, 1987). For example, Marsh and Bailey (1991), using 435 MTMM matrices based on real and simulated data showed that the CFA-CTCM model typically resulted in improper solutions (77% of the time) whereas the CFA-CTCU model nearly always (98% of the time) resulted in well-defined solutions. When both solutions were proper, parameter estimates based on the CFA-CTCU model tended to be more accurate and precise in relation to known parameter values based on simulated data. Even for data specifically constructed to have correlated method effects as posited in the CFA-CTCM model but not the CFA-CTCU model, the CFA-CTCU uniqueness model was more likely to converge to a proper solution and provided more accurate parameter estimates even though it was not able to completely able to fit the data, thus indicating that it was not a "true" model. Improper solutions for the CFA-CTUM and particularly the CFA-CTCM models were more likely when the MTMM design was small (i.e., 3Tx3M vs 5Tx5M), when the sample size was small, and when the assumption of unidimensional method effects was violated. From this practical perspective, the complications in comparing the CFA-CTCM, CFA-CTUM, and CFA-CTCU models may be of limited relevance because in many applications only the CFA-CTCU model results in a proper solution.

Covariance Component Analysis

Wothke (1984, 1987; also see Browne, 1989; Kenny & Kashy, in press) described the covariance component analysis (CCA) model that is based in part on earlier work by Bock (1960) and Bock and Bargmann (1966) and, in some ways, resembles the ANOVA approach discussed earlier. The "factors" in the CCA model are not based on freely estimated factor loadings as in the CFA approach, but are fixed contrast coefficients like those used in ANOVA. In fact, given the many parallels between the CCA and ANOVA models, it is curious that Wothke (1984, 1987) did not evaluate this earlier approach and its relation to his CCA model. The key parameter estimates in the CCA model are the relative size of variance components due to trait contrasts, method contrasts, and a general factor. In Wothke's parameterization of the CCA model, there is one general factor reflecting an average score across all the measures, T-1 trait contrast factors, and M-1 method contrast factors. According to the scale free version of the CCA model that is most appropriate for the analysis of MTMM data, the population covariance matrix Σ can be expressed as:

(1) $$\Sigma = D\,(K\,\Phi\,K')\,D + \Theta$$

where K is (M x T) x (M + T - 1) matrix of fixed othonormal column contrasts like those used in traditional ANOVA models, Φ is a (M + T -1) x (M + T - 1) variance-covariance matrix, Θ is a typically diagonal matrix of uniqueness terms, and D is a diagonal matrix of scaling constraints designed to absorb scaling constants so that the model can be fit to correlation matrices (Wothke, 1984, 1987).

The K matrix is a fixed set of coefficients constructed in the same way as in contrasts ANOVA. Thus, for example, for a 3T x 3M design with measured variables x(T1M1), x(T1M2) .. x(T3,M3), the 9 x 5 K matrix can be represented by:

(2)	0.333333	0.471405	0	0.471405	0
	0.333333	-0.235702	0.408248	0.471405	0
	0.333333	-0.235702	-0.408248	0.471405	0
	0.333333	0.471405	0	-0.235702	0.408248
	0.333333	-0.235702	0.408248	-0.235702	0.408248
	0.333333	-0.235702	-0.408248	-0.235702	0.408248
	0.333333	0.471405	0	-0.235702	-0.408248
	0.333333	-0.235702	0.408248	-0.235702	-0.408248
	0.333333	-0.235702	-0.408248	-0.235702	-0.408248

where (a) the first column of 9 coefficients reflects the general factor; it is like the "mean" contrast in a typical ANOVA; (b) the next two columns are the T-1 trait contrasts such that the first trait contrast represents the difference between T1 and the average of T2 and T3 and the second trait contrast reflects the different between T2 and T3 (averaged over methods); these are like the ANOVA contrasts used to reflect the T-1 degrees of freedom associated with T traits; and (c) the last two columns reflect the M-1 method contrasts such that the first method contrast represents the difference between the M1 and the average of M2 and M3 and the second method contrast reflects the difference between M2 and M3 (averaged over traits); these are like the ANOVA contrasts used to reflect the M-1 degrees of freedom associated with M traits. Whereas any alternative set of contrasts can be used (Browne, 1989; also see Kenny & Kashy, in press), Wothke (1984, 1987) argued that the use of orthonormal contrasts like in equation 2 facilitates subsequent interpretations.

The most important parameter estimates in the CCA model are in the variance/covariance matrix (Φ) reflecting the general factor, and the trait and method contrast factors. An arbitrainess of the scale is resolved by fixing the variance of the general factor to 1.0 so that variance estimates for the trait and method contrast factors are evaluated relative to the size of the general factor. In the "block diagonal" model considered here, covariances among trait contrast factors and among method contrast factors are estimated, but all other covariances are constrained to be zero. For the "scale free" version of the model that is most generally useful and appropriate for the analysis of correlations, Wothke (1987) noted that models with covariance terms involving the general factor are not identified when the MTMM matrix is small (i.e., 2T x 2M) and, apparently, are empirically underidentified for larger designs. Whereas it is possible to estimate correlations between traits and methods, Wothke's (1984, 1987) investigation with 23 MTMM matrices indicated that this model frequently resulted in improper solutions.

Wothke (1987) expanded his 1984 presentation by suggesting alternative summaries of the variance/covariance (Φ) matrix (e.g., generalized dispersion components, eigenstructures and associated eigenvalues, and varimax rotations) that may facilitate the interpretation of CCA parameter estimates. Nevertheless, the critical parameter estimates -- the variance components associated with the general factor and the trait and method contrast factors -- are not easily interpreted in relation to the terms typically used in MTMM studies. Making a related point, Browne (1989) indicated that relations among trait contrast factors and among method contrast factors provide only indirect information about correlations between traits and between methods, and that the arbitrainess of the contrasts used in K leads to an arbitrainess in the interpretation of the parameter estimates in Φ.

In an attempt to relate his model to traditional MTMM terminology, Wothke (1987; p. 38) proposed that convergent validity is supported if the variance/covariance matrix of method contrast factors approaches zero and that discriminant validly is established when the determinant of the variance/covariance matrix of trait contrast factors is large. There are, however, potential limitations with both these proposals.

1. Not even a complete absence of method effects provides support for convergent validity, whereas it is possible for convergent validity to exist even when there are substantial method effects. Hence, inferences based on method contrast factors are apparently a weak basis for inferring convergent validity.

2. A zero determinant of the trait contrast factor variance/covariance matrix indicates that at least one trait is a linear combination of the remaining traits, thereby precluding support for the discriminant validity of all the traits. However, the zero determinant could occur when all the traits are correlated 1.0 with each other (a complete lack of discriminant validity), or when all but one of the traits are uncorrelated but the one remaining trait is a linear combination of the other traits. In neither case would there be complete support for discriminant validity, but in the latter there apparently would be strong support for the discriminant validity of all but one of the traits. Hence, inferences about discriminant validity based on the determinant of the covariance matrix of trait factors may not be sufficiently sensitive to provide a useful indication of the extent of support for discriminant validity.

It is useful to examine similarities between the ANOVA and CCA models. Both models provide estimates of variance associated with a general factor, trait contrasts, and method contrasts. For both approaches:

(a) The CCA general factor, like the subjects term in the ANOVA model, is based on a subject's mean score across all measures so that the variance component is the between subject variance in this grand mean score. The relative size of this variance component reflects an overall average agreement across traits and methods; if scores reflecting different traits and methods are all equal within each subject but vary across subjects, then all the variance will be due to the general factor. Thus, in the ANOVA model, this term is defined as the convergent validity effect (although we noted that limitations with this interpretation).

(b) The variance components associated the trait contrast factors, like the trait x subject interaction in the ANOVA model, reflects the extent to which profiles of trait scores vary from subject to subject. To the extent that this component is large, subjects differ systematically in how they are ranked on the different traits. If the trait scores for each subject are equal, then the traits do not differentiate among subjects. Thus, this variance component provides an apparently useful indication of discriminant validity.

(c) The variance components associated with the method contrast factors, like the method x subject interaction in the ANOVA model, reflect the extent to which profiles of method scores vary from subject to subject. To the extent that this component is large, subjects differ systematically in how they are ranked on the different methods. If scores reflecting the different methods are equal for each subject, then the methods do not differentiate among subjects. Thus, this variance component provides an apparently useful indication of method effects.

(d) Both models emphasize the global evaluation of variance components and not the formative evaluation of specific measures, traits, and methods that apparently is an important contribution of the MTMM paradigm as a formative tool.

(e) The ANOVA and CCA models are sensitive to the orientation of the variables (i.e., changes in sign associated with a given variable). Thus, as noted earlier, the change in orientation of any variable will typically have a substantial effect on the variance components. (This characteristic of the CCA model and its implications are discussed further in relation to analyses of the Kelly and Fiske data.). As a consequence, it is probably advisable to reflect all traits so as to maximize the number and extent of positive correlations in the MTMM matrix.

There are, however, important differences in the ANOVA and CCA approaches. Inferences in the ANOVA model are based on measured variables whereas those in the CCA model are based on latent variables. In the ANOVA model the variance associated with the general factor is interpreted as an indication of convergent validity; Wothke placed little emphasis on the variance of the general factor in his CCA model except as a basis of comparison

for other variance components, but offered no criteria of convergent validity other than a lack of method effects. Noting limitations in Wothke's proposed test of convergent validity, we suggest that the size of the variance component associated with the general factor -- in relation to those associated with traits and methods -- may be the best indication of convergent validity available in the CCA model. There are also apparently important differences in how the variance components associated with each of the T-1 trait contrasts and the M-1 method contrasts are combined. In the ANOVA model, the variance components are combined additively, whereas Wothke's proposal to use the determinant implies a multiplicative combination.[1] The difference between the two approaches is clear in the example noted earlier in which all but one trait is uncorrelated with the others and the remaining trait is a linear combination of the others. An additive combination would result in a nonzero, possibly very large combined effect of the trait contrasts, whereas the product combination would result in a zero combined effect.

Composite Direct Product Model

The CFA and CCA models considered here implicitly assume that trait and method effects are additive. Observations by Campbell and O'Connell (1967, 1982) and others, however, suggest that the relation may be multiplicative or a combination of multiplicative and additive rather than strictly additive. Both the additive and multiplicative models posit that correlations between traits measured with the same method will be higher than correlations between traits measured with different methods -- a method effect. If this method effect is additive, then the increase in correlation due to this method effect is expected to be relatively similar for all correlations of differing magnitudes. Campbell and O'Connell, however, suggested that the method effects are systematically larger for traits that are more highly correlated and systematically smaller for traits that are less correlated. This empirical observation suggests that method effects have a multiplicative effect on trait correlations.

Campbell and O'Connell (1967, 1982) offered two different interpretations of this multiplicative effect. The differential augmentation perspective is that observed correlations are a multiplicative function of the true correlation and a method bias. According to this perspective, when true traits are uncorrelated there will be no bias (i.e., the method effect multiplied by zero is zero). In contrast, when traits that are substantially correlated the correlation between the traits based on the same method will be biased so long as the method effect is nonzero. This portrayal of method effects differs from the additive model that implicitly assumes that the size of method effects does not vary according the size of true trait correlations. The differential attenuation perspective suggests that the use of different methods will attenuate the true correlation between two traits. The extent of this attenuation, however, will vary according to the size of the correlation. If the true trait correlation is already zero, the correlation cannot be attenuated. In contrast, if the true trait correlation is substantial, then the empirical correlation can be attenuated substantially. According to this perspective, the correlation between two traits measured by the same method is the more accurate estimate of the true correlation, and this correlation is attenuated when different methods are used. This perspective is apparently consistent with the typical simplex pattern of relations observed in longitudinal data whereby the size of correlations between traits declines systematically as the time between the collection of the measures becomes longer.

Browne (1984), based in part on earlier work by Swain (1975), described the composite direct product (CDP) model that posits a multiplicative rather than an additive combination of trait and method effects. According to the CDP model there are two component correlation matrices in a MTMM matrix of correlations among latent variable scores (Pc), one containing correlations between latent traits (Pt) and the other containing correlations between latent methods (Pm). According to the CDP model, the covariance matrix of measured variables with dimension (mt x mt) can be expressed as:

(3) $\Sigma = D (Pm \times Pt + E) D$

where Pm is an (m x m) latent variable score correlation matrix of method components, Pt is a (t x t) latent variable score correlation matrix of trait components, D is a (mt x mt) positive definite, diagonal matrix of scale constraints reflecting latent variable score standard deviations, E is a positive, definite diagonal matrix of uniquenesses reflecting the ratio of unique score variance to latent variable score variance, and x indicates the right direct Kronecker product of Pm and Pt.

The values of D are typically of no interest and are designed primarily to absorb scaling changes such as those involved in going from a covariance matrix to a correlation matrix. The E values, however, represent the ratios of unique score standard deviations to latent variable score standard deviations. Browne (1984, 1989) noted that these values can be interpreted as the correlation between an observed and latent variable score, an "index of communality," when transformed by the formula:

(4) communality (Ti, Mr) = $1 / [1 + (E(Ti, Mr)]$

According to the CDP model, the correlation matrix Pc, appropriately corrected for attenuation, has the direct product structure:

$$Pc = Pm \times Pt$$

where Pm is the correlation matrix of relations among latent method factors with a typical element being r(Mr, Ms) and Pt is the correlation matrix of relations among latent trait factors with a typical element being r(Ti, Tj). From this definition it follows that for latent variable scores

(4) r(TiMr, TjMs) = r(Ti,Tj) r(Mr,Ms)

It is useful to demonstrate the relation between Pm, Pt, and Pc using, for example, a 2T x 3M design.

(5)

Pm =			x	Pt =		Pc =	Pc	x	Pt	=					
1				1		1									
M21	1			T21	1	T21		1							
M31	M32	1				M21		T21xM21		1					
						T21xM21		M21		T21		1			
						M31		T21xM31		M32		T21xM32		1	
						T21xM31		M31		T21xM32		M32		T21	1

where, for example, T21 is the correlation between traits 1 and 2 and T21xM31 is the product of the correlation between traits 1 and 2 and the correlation between methods 3 and 1. All elements of Pt are multiplied by each element of Pm. Thus, the relation between traits 1 and 2 measured with method 1 is T21 multiplied by M11 = 1 so that the product is simply T21. Similarly, the relation between trait 1 measured with methods 1 and 2 (i.e., a convergent validity) is M21 times T11 = 1 so that the convergent validity is simply M21. Thus, the coefficients in the off-diagonal of Pm reflect convergent validity. Note also, that the correlation between the same traits is assumed to be constant across all methods (i.e., r(T1M1,T2M1) = r(T1M2,T2M2) = r(T1M3,T2M3) = T21). Similarly, the correlation between two methods -- convergent validity -- is assumed to be the same across all traits (i.e., r(T1M1,T1M2) = r(T2M1,T2M2) = M21). Because the 15 off-diagonal values in this Pc are expressed in terms of only 4 estimated parameters (T21, M21, M31, M32), the CDP model is very parsimonious.

Browne (1984, 1989; also see Bagozzi & Yi, 1990; Cudeck, 1988) notes that an important advantage of this model is that it provides parameter estimates that can be used to evaluate the original 4 Campbell-Fiske guidelines.

1 r(TiMp, TiMq) = r(Ti,Ti) r(Mp,Mq) = r(Mp,Mq) >> 0.

According to the CDP model, each convergent validity for latent variable scores is equal to one of the off-diagonal values in Pm. Hence, the first Campbell-Fiske criterion is satisfied whenever all the off-diagonal values in Pm are statistically significant, large, and positive.

2 r(TiMp, TiMq) > r(TiMp, TjMq) implies

[r(TiMp, TjMq)] / [r(TiMp, TiMq)] = [r(Ti,Tj) r(Mp,Mq)] / [r(Mp, Mq)] = r(Ti, Tj) < 1.0

According to the CDP model, the latent variable trait correlations, the off-diagonal values in Pt, are the ratio of HTHM correlations to the convergent validities. Hence, the second Campbell-Fiske criterion is met whenever the off-diagonal values of Pt are less than 1.0. This will always be the case so long as the CDP solution is proper such that Pt is positive definite.

3 r(TiMp, TiMq) > r(TiMp, TjMp) implies

[r(TiMp, TjMp)] / [r(TiMp, TiMq)] = [r(Ti,Tj)] / [r(Mp, Mq)] < 1.0

According to the CDP model, the ratio of HTMM correlations to the convergent validities is the ratio of trait correlations to method correlations. Hence the third Campbell-Fiske guideline is met when all the off-diagonal values in Pt are less than all the off-diagonal values in Pm.

4 r(TiMp, TjMq) > r(TkMp, TlMq) implies r(TiMr, TjMs) > r(TkMr, TlMs)

This criterion is met whenever the CDP model fits the data because:

r(TiMr, TjMs) / r(TkMr, TlMs) = r(Ti, Tj)/ r(Tk, Tl) has the same value for any Mr or Ms.

Although not explicitly noted in previous presentations of the CDP model, it is also possible to interpret the additional guidelines from the Campbell-Fiske approach (see guidelines 5 and 6 discussed earlier) in terms of the CDP model.

5 $r (TiMp, TjMp)/ [(r (TiMp, TiMp) r(TjMp, TjMp)]^{1/2} << 1$

Because the values in Pt reflect correlations among latent trait factors, this condition is satisfied whenever the CDP model results in a proper solution in which Pt is positive definite. Also, as noted earlier, the CDP model provides an estimated communality the can serve as an estimate of reliability.

6. r(TiMr, TjMr) > r(TiMr, TjMs) implies [r(TiMr, TjMs)] / [r(TiMr, TjMr)] =

[r(Ti,Tj) r(Mr, Ms)] / [r(Ti,Tj) r(Mr,Mr)] = [r(Ti,Tj) r(Mr, Ms)]/ [r(Ti,Tj)] = r(Mr, Ms) < 1.0

According to the CDP model, there are method effects whenever the correlations in Pm are less than 1, and so there are always method effects when the CDP results in a proper solution in which Pm is positive definite. Also evident in this derivation is the observation that r(Mr, Ms) reflects both convergent validity (see guideline 1) and method effects (i.e., the ratio of HTHM and the corresponding HTMM correlations). Whereas this observation appears paradoxical from the traditional "additive" perspective, it follows naturally from the "multiplicative" perspective underlying the CDP model.

It is also possible to place further constraints on the CDP model that may be useful in particular situations. Thus, for example, it is possible to further restrict the structure of E, the diagonal matrix of uniquenesses, so that it also has a direct product structure (Browne, 1984, 1989; Wothke & Browne, 1989). Also, if the covariance matrix rather than the correlation matrix is analyzed, it is possible to further restrict the structure of D, the diagonal matrix of scale constraints reflecting latent variable score standard deviations. Such models may be useful when the MTMM data reflects multiple battery data, such as when the same measures are collected on multiple occasions, but are apparently less relevant to other MTMM designs and are not central to interpretations of convergent validity, discriminant validity, and method effects (for further information see Browne, 1984; 1989).

Results of the Latent Construct Approaches

Proper Solutions and Goodness of Fit.

Initially we focus on the ability of the latent variable models -- the CFA, CCA, and CDP models -- to fit the data. The evaluation of fit in covariance structure analysis has recently received considerable attention and a detailed discussion of the issues is beyond the scope of this study (see Bentler, 1990; Cudeck & Henly, 1991; Marsh, Balla, and McDonald, 1988; McDonald and Marsh, 1990 for general discussions and Marsh, 1989, for a discussion in relation to MTMM data). Whereas there are no well established guidelines for what minimal conditions constitute an adequate fit, a general approach is to: (a) establish that the solution is "proper" by establishing that the model is identified, the iterative estimation procedure converges, parameter estimates are within the range of permissible values (i.e., are inside the admissile parameter space), and the size of the standard error of each parameter estimate is reasonable; (b) examine the parameter estimates in relation to the substantive, a priori model and common sense; (c) evaluate the X^2 and subjective indices of fit for the model and compare these to values obtained from alternative models.

In the evaluation of MTMM models there is an unfortunate tendency to deemphasize the first two points. If a solution is ill-defined, then further interpretations must be made cautiously if at all. If the parameter estimates make no sense in relation to the substantive, a priori model, then fit may be irrelevant. For example, if two indicators of the same trait factor are supposed to load in the same direction but actually load in the opposite direction, then the results do not support the construct validity of the trait even if the model fits the data well. In this respect, the first criterion is a prerequisite for the next two and the second criterion is a prerequisite for the third.

For each of the latent variable models, solutions are proper if the model is identified and if the estimated parameters fall within their permissible range. For models considered here a proper solution requires that all estimated covariance matrices should be positive definite. In the CFA models this means that there are no negative or zero variance estimates and that factor correlations do not exceed 1.0. For the CDP and CCA models this means that the matrices of scaling components and error components contain no negative or zero values. Using reparameterizations such as those suggested by Rindskopf (1983; also see Marsh, 1989) it is possible to restrict, for example, a negative variance estimate to be non-negative. Typically this results in the offending parameter taking on a zero value that is on the boundary of the permissible parameter space and in a slight decrement in goodness of fit reflecting this implicit inequality constraint. Marsh (1989) argued that whereas this may be useful in some situations, it is important to emphasize that a solution with, for example, a zero variance estimate is still improper and should be treated with the same caution as if the parameter estimate were negative. In this sense, the reparameterization does not alter the underlying problem but merely serves to make it less obvious. Making a similar point, Joreskog and Sorbom (1989) emphatically stated that "it should be emphasized that constraining error variances to be non-negative does not really solve the problem. Zero estimates of error variances are as unacceptable as are negative estimates" (p. 215). There is an ongoing debate about whether improper solutions warrant any serious consideration and, if they are considered, the conditions under which interpretations are justified. Not wanting to enter this debate in relation to particular applications in the present investigation, our position is that if a model frequently results in improper solutions across a wide range of applications for which the model is intended, then the usefulness of the model is limited (see Marsh & Bailey, 1991).

Goodness of fit is evaluated in part with an overall X^2 test. As typically employed the posited model is rejected if the X^2 is large relative to the degrees-of-freedom (df), and accepted if the X^2 is small and nonsignificant. However, hypothesized models such as those considered here are best regarded as approximations to reality rather than exact statements of truth so that any model can be rejected if the sample size is sufficiently large. Conversely,

almost any model will be "accepted" if the sample size is sufficiently small. From this perspective Cudeck and Browne (1983) and many others have argued that it is preferable to depart from the hypothesis testing approach that assumes that any model will exactly fit the data.

As emphasized by Bentler (1990), when two models are nested the statistical significance of the difference in the X^2s can be tested relative to the difference in their df. Widaman (1985) emphasized this feature in developing his taxonomy of MTMM models and in comparing the fit of different models. However, the problems associated with hypothesis testing based on the X^2 statistic also apply to tests of X^2 differences. Furthermore, many important comparisons are not nested and so cannot be compared using this procedure. For example, whereas the CFA-CTUM is nested under both the CFA-CTCM and CFA-CTCU models, neither of these latter two models is nested under the other. Nevertheless, a pattern of nested relations does facilitate interpretations in differences in fit.

Researchers have developed a plethora of different indices of fit, but there is no clear consensus about which are the most useful. Whereas a comparison of different indices is beyond the scope of this study, we present results for the: the X^2 that can be used to compute values for most other indices; the relative noncentrality index (RNI; McDonald & Marsh, 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973; also see Marsh, Balla & McDonald, 1988; McDonald & Marsh, 1990), and the single-sample cross-validation index (Ck; Browne & Cudeck, 1989; Cudeck & Henly, 1991). Both the TLI and RNI indices scale goodness of fit along a scale that, except for sampling fluctuations, varies between 0 and 1. Values greater than .9 are typically interpreted as indicating an acceptable fit, although it may be more useful to compare the values of alternative models. The TLI and RNI differ in that the TLI contains a penalty function based on the number of estimated parameters whereas the RNI does not. The Ck index is designed to select the model that will cross-validate most effectively, and so it imposes a penalty that is a an increasing function of the number of estimated parameters and a decreasing function of the sample size.

The minimal condition for an acceptable fit is a proper solution. If the solution is improper, then further consideration should be pursued with extreme caution and may be dubious. This problem has been prevalent in the application of the CFA models -- particularly the CFA-CTCM model. The prevalence of this problem led, in part, to recommendations for the CFA-CTCU, CCA, and CDP models.

The CFA-CTCU, CFA-CT and CFA-CTUM models resulted in 0, 1 and 2 improper solutions respectively, whereas the CCA and CDP models each resulted in one improper solution. Consistent with previous research (e.g., Marsh, 1989; Marsh & Bailey, 1991; Wothke, 1984; 1987), the CFA-CTCM resulted in a proper solution for only 1 of the 5 MTMM matrices (Table 2). All other models considered here performed better than the CFA-CTCM model in terms of resulting in proper solutions. Consistent with findings by Marsh and Bailey (1991), the one proper solution for the CFA-CTCM model was obtained when the sample size (N = 1200) and MTMM design were large (5Tx3M vs. 3Tx3M). The consistency with which the CFA-CTCM model results in improper solutions undermines its usefulness and suggests, perhaps, that it should not be given a central role in the empirical evaluation of MTMM data. This is a very serious problem because most applications of the CFA approach -- and the relatively few comparisons of the CFA approach with other latent variable approaches -- have relied exclusively or primarily on the CFA-CTCM model.

Insert Table 2 About Here

For all five data sets, the X^2s associated with the CFA-CTCU model were better than those for any the other models that resulted in proper solutions. The TLI, incorporating a penalty for a lack of parsimony, was marginally better for the CFA-CTUM model than the CFA-CTCU model for the Kelly and Fiske data, but the TLI was better for the CFA-CTCU model than any other model that resulted in a proper solution for each of the other data sets. The Ck

that imposes a penalty function that depends on sample size, was better for CDP model than the CFA-CTCU model for the Lawler data that had the smallest sample size. The TLIs and RNIs were all substantially greater than .9 for the CFA-CTCU model each of the five data sets.

The examination of goodness of fit -- both the number of improper solutions and the fit indices -- provide support for the CFA-CTCU model. There are, however, some relevant qualifications to these conclusions. For all the data sets, several different models provided apparently acceptable fits in that the solutions were proper and both the TLI and RNI were larger than .9. Because the CFA-CTCU model is considerably less parsimonious -- uses more estimated parameters to fit the same data -- it may be premature to claim that it fits the data better. Also, because the CFA, CCA and CDP models are so different, it is important to evaluate the usefulness of alternative models in terms of interpretations of the parameter estimates in relation to providing information about convergent validity, discriminant validity, and method effects and providing a formative evaluation of each trait-method unit.

The CFA Models: Goodness of Fit

The comparison of the fit indices for the various CFA models (Table 2) is facilitated by the nesting relations among the models. The strategies used to compare these model outlined here appear to offer a reasonable basis for evaluating assumptions underlying the models. The CFA-CT model is nested under the other CFA models considered here. The size of the difference in fit between the CFA-CT model and each of the other models provides an indication of the size of the method effects. For all five data sets, the fit of the CFA-CT model is significantly poorer than the other CFA models, indicating the existence of method effects. Whereas the comparisons vary somewhat depending on which models are compared, the inferred method effects are smaller for the Byrne data, the YIT data, and to a lesser extent, the Kelly and Fiske data. In contrast, the size of method effects are larger for the Lawler data and the Ireland data (although the improper solution for the CFA-CT model for the Ireland data dictates caution).

The comparison of the CFA-CTUM and CFA-CTCU models provides a test of the unidimensionality of method effects associated with each method when T > 3 and both models result in proper solutions. For the two data sets with T=3 (the Byrne data and the Lawler data) the CFA-CTUM and CFA-CTCU models are equivalent so long as both result in proper solutions, but the CFA-CTUM solutions were both improper (Table 2). For two of three remaining data sets with T=5, the fit of the CFA-CTCU is significantly better than the CFA-CTUM model, suggesting that the method effects are not unidimensional. For one data set (Kelly and Fiske) the CFA-CTUM and CFA-CTCU do not differ significantly, suggesting that the method effects in this study are unidimensional.

The comparison of the CFA-CTUM and CFA-CTCM models provides a test of whether the method effects associated with different methods are correlated, so long as both models result in proper solutions. The fit of the CFA-CTCM is consistently better than the CFA-CTUM model, suggesting that effects associated with different methods may be correlated. These results must, however, be viewed cautiously since the CFA-CTCM resulted in improper solutions for all but the YIT data. For the YIT data, the fit of the CFA-CTCM model is better than the CFA-CTUM model, but the difference in fit (e.g., TLIs of .978 and .982) is very small.

In general, the CFA-CTCM and CFA-CTCU models are not nested. When T=3, however, the CFA-CTCU and CFA-CTUM models are equivalent (method effects are necessarily unidimensional) and so the CFA-CTCU model is nested under the CFA-CTCM model. For the two data sets with T=3, the fit of the CFA-CTCM fit is marginally better than that of the CFA-CTCU. Interpretations must be made cautiously, however, since the CFA-CTCM solutions are improper. For the three data sets with T=5, the CFA-CTCM and CFA-CTCU models are not nested; the CFA-CTCM model fit better in one case whereas the CFA-CTCU fit better in the other two cases. Except

for the YIT data in which the CFA-CTCU model fit better, however, the improper solutions for the CFA-CTCM model dictate caution in these comparisons.

Comparisons among the CFA models are most useful for the YIT data since all the CFA models resulted in proper solutions. For this data set, most of the variance can be explained by the CFA-CT model (TLI = .903), although models with method effects (TLIs of .978 - .994) fit the data significantly better. The CFA-CTUM/CFA-CTCM comparison (.978 vs. .982) suggests that the effects associated with different methods are slightly correlated. The CFA-CTUM/CFA-CTCU comparison (.978 vs. .994) suggests that the various effects associated with each method are not unidimensional. Overall the CFA-CTCU model fits the best, even though there is some indication that its assumption of uncorrelated method effects is violated to a small extent (as evidenced by the CFA-CTUM/CFA-CTCM comparison). Hence, it may be useful to compare parameter estimates for the CFA-CTCM and CFA-CTCU models (see below).

The CFA Models: Interpretation of Parameter Estimates.

Parameter estimates for the CFA-CTCU model are summarized for all five data sets in Appendix 2. Also presented is the CFA-CTUM solution for Kelly and Fiske data that did not differ significantly from the CFA-CTCU solution, and the CFA-CTCM solution for the YIT data that was the only case in which this model resulted in a proper solution. For all models, large and statistically significance trait factor loadings provide an indication of convergent validity whereas large trait factor correlations -- particularly those approaching 1.0 -- suggest a lack of discriminant validity. Method effects are inferred from large and statistically significant method factor loadings in the CFA-CTCM and CFA-CTUM models, and from large and statistically significant correlated uniquenesses (among different variables assessed by the same method) in the CFA-CTCU model.

Byrne data. In the CFA-CTCU solution (Appendix 2) the trait factor loadings are consistently very large, the trait factor correlations are small or moderate, and the correlated uniqueness are small to moderate. As predicted, correlations between T2 and T3 are close to zero whereas other trait correlations are larger. It is also evident that method effects are smaller for M1 than for M2 and particularly M3, whereas trait effects are smaller for M3. These results provide strong support for the construct validity of interpretations of these data.

Lawler data. In the CFA-CTCU solution (Appendix 2) the trait factor loadings are large for M1 and M2, but small or nonsignificant for M3. The trait correlations are moderately large, but do not approach 1.0. Correlated uniquenesses are small to moderate. These results provide reasonably strong support for the construct validity of interpretations of measures associated with M1 and M2, but may call into question those based on M3 where T3M3 is the only variable with a significant trait factor loading.

YIT data. Solutions are presented (Appendix 2) for both the CFA-CTCU and CFA-CTCM methods since this is the only data set in which the CFA-CTCM model resulted in a proper solution. For both models, the trait factor loadings are consistently high whereas the trait factor correlations are small to moderate. Although these parameter estimates are similar in the two models, there is a tendency for trait factor loadings and trait factor correlations to be somewhat higher in the CFA-CFCU model (also see Marsh and Bailey, 1991; Kenny & Kashy, in press). Other parameter estimates in the two models, however, are not so easily compared. Correlated uniquenesses in the CFA-CTCU model tend to be small and more than half are nonsignificant, indicating weak method effects. Method factor loadings in the CFA-CTCM model are small to moderate but most are statistically significant, apparently providing somewhat stronger evidence of method effects than the CFA-CTCU model. In the CFA-CTCM model, the M1/M2 and M1/M3 correlations are small, but the M2/M3 correlation is moderate. In the CFA-CTCU model, effects associated with different methods are assumed to be uncorrelated. The uniqueness terms in the CFA-

CTCU model are systematically larger than those in the CFA-CTCM model because they include the effects of both the uniqueness and method effects in the CFA-CTCM model. For the same reason, the squared multiple correlations (SMCs) are smaller for the CFA-CTCU model than the CFA-CTCM model. If the effects associated with a single method are unidimensional, the CFA-CTCM model provides a more parsimonious and useful representation of method effects (i.e., the squared method factor loading can be interpreted as the proportion of variance due to method effects). If, on the other hand, the method effects are not unidimensional, then this convenient summary offered by the CFA-CTCM may be inappropriate.

Kelly and Fiske data. Here the CFA-CTCU and CFA-CTUM solutions (Appendix 2) are compared. Given that the two solutions are nested and do not differ significantly, it is not surprising that the trait factor loadings and trait factor correlations are similar. Trait factor loadings are consistently large for M1, large for all but T4M2 for M2, and moderate for M3. Trait correlations are small to moderate. For both models method effects are small for M1 (except T3M1) and small to moderate for M2 and M3. Particularly because the difference between the two models is nonsignificant, the more parsimonious, convenient representation of method effects in the CFA-CTUM model is preferable to the CFA-CTCU model in this example.

Ireland data. Here, the CFA-CTCU and CFA-CTUM solutions (Appendix 2) are compared. Whereas the CFA-CTCU model fit the data significantly better, the difference was not large (TLIs of .988 and .964). Again, the trait factor loadings trait factor correlations are very similar for the two models. The trait factor loadings are consistently high, indicating convergent validity, but the trait factor correlations are so high that there is little or no support for divergent validity. Although the method factor loadings and correlated uniquenesses are not directly comparable, both indicate moderate to large method effects. These results thus suggest a good overall agreement across the different methods, but a clear lack of discriminant validity.

The CCA Model

CCA parameter estimates for the five data sets are summarised in Appendix 3. The critical parameter estimates are the variance components associated with the trait and method contrast factors. For present purposes we interpret large variance components associated with trait contrast factors as support for discriminant validity and large variance components associated with method contrast factors as evidence of method effects. With misgivings based on limitations noted earlier, we interpret large variance components associated with the general factor -- compared to those associated with trait and method contrasts -- as support for convergent validity because this is apparently the only available indicator of convergent validity. Also, because the general variance component is fixed to 1.0 to establish the scale of the other variance components, the size of all other variance components must be interpreted in relation to that of the general factor.

Byrne data. The variance components associated with trait contrasts are larger than those associated with method contrasts (see Appendix 3), but the largest component is for the general factor. This suggests the existence of weak method effects, clear support for discriminant validity, and even stronger support convergent validity.

Lawler data. The variance components associated with method contrasts are larger than those associated with trait contrasts (Appendix 3), but the largest component is for the general factor. This suggests substantial method effects, limited support for discriminant validity, and strong support for what is interpreted to be convergent validity.

YIT data. The variance components associated with method contrasts are smaller than those associated with trait contrasts (Appendix 3), but the largest component is for the general factor. This suggests a relative lack of method effects, clear support for discriminant validity, and strong support for what is interpreted to be convergent validity.

Kelly and Fiske data. The variance components associated with method contrasts are generally smaller than those associated with trait contrasts (Appendix 3), but the component for the general factor is larger than all but one of the components associated with traits and all the components associated with methods. This suggests relatively small method effects, support for discriminant validity, and support for what is interpreted to be convergent validity.

A potential weakness of the CCA approach noted earlier is its sensitivity to the orientation of traits that is evident in a second analysis of the Kelly and Fiske data. T3, ratings of the trait "seriousness," tends to be negatively correlated with the other traits (see Appendix 1). We reanalyzed the Kelly and Fiske data after reflecting the orientation of T3 (i.e., reversing all the signs of correlations associated with T3M1, T3M2, and T3M3). This resulted in a different chi-square (139.09 vs. 115.87) and substantially different variance components (Appendix 3). Specifically, the variance components associated with the trait and method contrasts are substantially smaller in the reanalysis. This follows because the variance component of the general factor is the between subject variance on the mean score and reflects the average correlation among all the (latent) measures. Its value is fixed at 1.0 and the size of other variance components are scaled in relation to its value. By reversing the signs of the predominantly negative correlations between T3 indicators and the other measures, the average correlation among measures is increased as is the between subject variance on the mean score. This results in a higher proportion of the variance due to the general factor, which in the CCA model is translated into lower variance components due to trait and method contrast factors. It should be noted that if all measures in a MTMM study are substantially and positively correlated, reversing the orientation of one of the traits would typically have even larger effects than in the Kelly and Fiske data. The reflection of "negatively oriented" traits so as to maximize the average correlation among all traits is probably a reasonable rule to overcome this apparent arbitrainess in the CCA approach, although Wothke (1987) did not do this with the Kelly and Fiske data. More generally, however, the extreme sensitivity of the CCA approach to the orientation of traits appears to be a potentially serious limitation in the approach.

Ireland data. The CCA model resulted in an improper solution. As a pragmatic alternative, we fit the completely diagonal version of the CCA in which covariances among trait contrast factors and among method contrast factors were all fixed to be zero (see CCA-Diag in Appendix 3). The completely diagonal model is not generally recommended because it depends on the appropriateness of the particular contrasts in a way that is idiosyncratic to a particular application (see Wothke, 1984; 1987). The variance components associated with method contrasts are larger than those associated with trait contrasts, but the component for the general factor is much larger than those associated with either trait or method contrasts. This suggests small method effects, almost no support for discriminant validity, and strong support for what is interpreted to be convergent validity.

The CDP Model

CDP parameter estimates for the five data sets are summarised in Appendix 4. The critical parameter estimates are the correlations among trait factors and among method factors. As noted earlier: (a) high method factor correlations are interpreted as support for convergent validity (agreement between measures based on different methods); (b) trait factors substantially smaller than 1.0 and smaller than the method factor correlations are interpreted as support for discriminant validity. Method effects have a very different interpretation within the context of the "multiplicative" CDP model than in the additive models considered earlier and, according to the CDP model, there are always method effects whenever the CDP solution is proper.

Byrne data. The Pm correlations for the Byrne data are consistently very large and consistently larger than the Pt correlations, whereas the Pt correlations are consistently smaller than 1. This implies clear support for all the Campbell-Fiske guidelines and strong support for the construct validity of these measures. The relative lack of

correlation between T2 and T3 observed in the MTMM matrix is evident in Pt. Similarly, the apparently stronger agreement between measures based on M1 and M2 is evident in Pm.

Lawler data. The Pm correlations in the Lawler data are all statistically significant, but only the M1M2 correlation is substantial. The Pm correlations, except for the M1M2 correlation, are consistently smaller than the Pt correlations. In general, these results suggest modest support for convergent validity and a lack of discriminant validity, although there is support for the convergent validity of measures based on M1 and M2. Thus, the apparently better agreement between measures based on M1 and M2 observed in the MTMM is also apparent in Pm.

YIT data. The Pm correlations for the YIT data are consistently large and consistently larger than the Pt correlations, whereas the Pt correlations are small to moderate. This implies clear support for all the Campbell-Fiske guidelines and strong support for the construct validity of these measures. The better agreement between measures based on M2 and M3 observed in the MTMM is also apparent in Pm. Whereas correlations among traits are not large, the patterns of differences in the MTMM matrix (Appendix 1) are evident in Pt.

Kelly and Fiske data. The Pm correlations for the Kelly and Fiske data are moderate to large. Whereas the Pt correlations are consistently less than 1.0, some are larger than the Pm correlations. Whereas the Pm correlation between M1 and M2 is consistently larger than the Pt correlations, the other correlations in Pm are not. These results suggest clear support for convergent validity, but only weak support for discriminant validity. The better agreement between measures based on M2 and M3 observed in the MTMM matrix (Appendix 1) is apparent in Pm. Similarly, the pattern of correlations among traits in the MTMM matrix is evident in Pt.

Ireland data. The Pm correlations for the Ireland are consistently large. The Pt correlations, however, are consistently even larger and often approach 1.0. These results suggest clear support for convergent validity, but no support for discriminant validity.

Discussion and Recommendations

Five approaches to the analysis of MTMM data are described here. Even though all the approaches use a similar terminology (convergent validity, discriminant validity, and method effects), they employ different operationalizations of these terms and so are not equivalent. This has led to considerable confusion in MTMM research. For this reason it is useful to summarize strengths and weakness of the different approaches and to offer recommendations for their use.

The Campbell-Fiske approach continue to be the best known and most widely applied of the approaches. Despite important limitations such as a reliance on measured variables instead of latent constructs, this approach continues to be a potentially useful and heuristic approach to the formative evaluation of MTMM data. This approach is also the basis, to a greater or lesser extent, of subsequent approaches. For this reason we recommend that a systematic application of the expanded set of Campbell-Fiske guidelines to provide a preliminary inspection of the MTMM data prior to the application of more sophisticated approaches. Consistent with the Campbell and Fiske's recommendations and the many limitations in this approach, it should be used as a formative evaluation of the data that focuses on specific trait-method units and not a global summative statement. The guidelines should not be the sole basis for evaluating MTMM data.

The reliance of each of alternative approaches on the original Campbell-Fiske approach has both advantages and limitations. The widely known terminology used in the Campbell and Fiske approach has provided an important starting point for other approaches. Nevertheless, the terms convergent validity, discriminant validity, and method effects were not adequately defined in the Campbell-Fiske approach and there is considerable ambiguity in how their guidelines relate to these different aspects of MTMM data. Partly as a consequence of this initial ambiguity,

subsequent approaches have each adapted somewhat different and possibly incompatible guidelines for these different characteristics. As asserted by Kenny and Kashy (in press), it appears that Campbell and Fiske (1959) implicitly based their original guidelines on a general CFA model. In the CFA-CTCM model it is clear that convergent validity, discriminant validity, and method effects are a function of the sizes of trait factor loadings, trait factor correlations, and method factor loadings respectively. Because of this apparently unambiguous interpretation of these features based on the CFA-CTCM model, we recommend that this model should be used as a touchstone for defining terminology in MTMM studies and for evaluating new models or different approaches. The fact that the CFA-CTCM model typically results in improper or unstable solutions means that other approaches are needed. Similarly, lamenting that "the rich detail of the general CFA model is not a realistically achievable goal" (p. 22), Kenny and Kashy argued that it is necessary to introduce simplifying conditions to achieve generally interpretable results.

Recommendations For Alternative Approaches.

The ANOVA approach provides convenient summative statistics about the relative size of convergent validity, discriminant validity, and method effects. There are, however, important limitations that apparently undermine its usefulness. The effects in the ANOVA model bear only a tangential relation to the typical meaning of discriminant validity, method effects, and particularly convergent validity. Also, this approach offers very little formative information about the effectiveness of particular traits, methods, or trait-method units. A serious limitation to the ANOVA model is that, like the Campbell-Fiske guidelines, it is based upon inferences about measured variables instead of latent traits. Because whatever advantages there are to this approach are apparently served more effectively by the CCA model, the ANOVA approach is not recommended.

The CFA approach is the most widely used latent variable approach to the evaluation of MTMM data. The comparison of different models and the comparison of parameter estimates in models reflecting trait and method effects provides clear evidence about convergent validity, discriminant validity, and method effects. A major limitation of this approach has been its reliance on the CFA-CTCM model that typically results in improper solutions. Furthermore, even when the CFA-CTCM does result in a technically proper solution, the solution may be sufficiently unstable that parameter estimates should be evaluated cautiously in relation to potentially large standard errors. Results summarized here, consistent with a large body of additional research (e.g., Marsh, 1989; Marsh & Bailey, 1991; Kenny & Kashy, in press), indicates that the problem of improper and unstable solutions is largely overcome through the application of the CFA-CTCU model. We recommend that at least the subset of CFA models considered here should be applied in all MTMM studies, but that the major emphasis should be placed on only those models that result in proper solutions. The preferred model within this set will depend on which models result in proper solutions and ability of the alternative models to fit the data, but a growing body of experience suggests that the CFA-CTCU model is the strongest model in the CFA approach.

The CCA approach, like the ANOVA approach, provides convenient summative statistics for effects that we have interpreted to correspond to reflect convergent validity, discriminant validly, and method effects. The important advantage of the CCA approach over the ANOVA approach is that inferences are based on relations among latent variables instead of measured variables. Nevertheless, other problems identified with the ANOVA model are also evident in the CCA model. These include an apparent ambiguity in how CCA parameter estimates relate to terminology typically used in MTMM studies, a lack of formative information about the performance of specific traits, methods, and trait-method units, and a sensitivity to the orientation of the traits. In addition, there is apparently no clear resolution on how best to combine the variance components associated with trait contrasts and those associated with method contrasts. Because of these apparent limitations, we do not recommend the routine

application of the CCA model for general use. It is possible, however, that further development of the approach along the lines proposed by Wothke (1987) and by Kenny and Kashy (in press) may overcome these limitations and provide a more generally useful approach.

The CDP model offers a mathematically elegant and parsimonious model of MTMM data. Whereas it has not been applied as widely as other latent variable approaches -- particularly the CFA approach -- results summarized here and those described in earlier research (e.g., Bagozzi & Yi, 1990; Browne, 1984, 1989; Cudeck, 1989) suggest that it typically results in proper solutions. Consistent with Browne's claim, the CDP model provides clear evidence about the Campbell-Fiske guidelines and about convergent and discriminant validity as embodied in these guidelines. The CDP model also provides parameter estimates that are typically consistent with those observed in the MTMM matrices. Therefore, subject to the continued demonstration of its success, we recommend that the CDP model should be used in MTMM studies.

Even though we endorse the continued use of the CDP model, we do so with some misgiving. Its parsimony is achieved at the expense of implicit assumptions that we find worrisome such as: (a) the convergent validities for all the different traits are equal (i.e., r(TiM1, TiM2) = r(M1, M2) for all values of i); (b) the size of method effects is the same for different traits (i.e., r(TiMr,TjMs)/r(TiMr,TjMr) = r(Ti,Tj) x r(Mr,Ms)/ r(Ti,Tj) = r (Mr,Ms) for all values of i and j); and (c) the size of correlations among traits is the same for all methods (i.e., r(TiMr,TjMr) = r(Ti,Tj) for all values of r). Pt correlations typically reflect the pattern of correlations among traits in the MTMM matrix, but only if this pattern is consistent across methods. Pm correlations typically reflect the extent of agreement between different methods, but only if the agreement is consistent across all traits. Whereas the overall fit of the model provides an indirect test of these assumptions, common sense suggests that they will typically be false so that a more detailed evaluation of the implications of violating these assumptions is needed in actual applications of the CDP model. Also, because of these implicit invariance constraints, the CDP model does not provide a very useful formative evaluation of specific trait-method units.

We also have some broader, philosophical concerns about the CDP model. The model, at least as applied to MTMM data, is apparently based on an uncritical acceptance of the original Campbell-Fiske guidelines. Thus, for example, Browne (1989) noted that "Campbell & Fiske (1959) listed four requirements for multitrait-multimethod correlation matrices that have become generally accepted. We shall be concerned with the investigation of these requirements" (p. xx). Whereas we agree that the heuristic value and intent of the Campbell-Fiske guidelines is widely endorsed, we do not concur that their literal translation as "requirements" as embodied in the CDP model is widely accepted. Indeed, it is the many problems and potential ambiguities in the guidelines that has spawned so many alternative approaches. Whereas the application of the CDP approach certainly provides an objectivity to evaluating the Campbell-Fiske guidelines, it is not clear that the CDP model eliminates widely recognized ambiguities in the interpretation of the Campbell-Fiske guidelines. Furthermore, if the underlying assumption of a multiplicative relation between traits and methods is taken literally, then the logic of the Campbell-Fiske guidelines and even the logic of the classical approach to test theory appears to be problematic. Whereas Browne (1984) has not claimed that support the CDP model necessarily leads to such dire consequences, we nevertheless find paradoxical the assumption that support for the CDP implies a multiplicative relation between latent traits and latent methods and provides a basis for evaluating the Campbell-Fiske guidelines that appear to be bases on an assumption of additivity that is invalidated by this multiplicative relationship. More generally, we are loath to relinquish the many conceptual and theoretical advantages in the additive assumption of variance components explicit in classical test theory and

conventional factor analysis that would have to be abandoned if such a multiplicative model were taken literally (see Campbell & O'Connell, 1967; 1982).

Comparison of the CFA and CDP Approaches

We have recommended the continued use of the CFA and CDP models, and so it is relevant to contrast the two approaches. Both the CDP and at least the CFA-CTCU models typically result in proper solutions. Consistent with Bagozzi and Yi (1990)[2], we found that CFA models fit real data better than the CDP model. Previous research, however, should be evaluated cautiously because all prior comparisons of the CDP and CFA approaches apparently were based on the CFA-CTCM that is known to be prone to improper and unstable solutions. Thus, for example, the improper CFA-CTCM solution with the Kelly-Fiske data and with the Lawler data have been used to argue for the superiority of both the CCA and CDP approaches over the CFA approach, but the CFA-CTCU solution is proper for both these examples. Nevertheless, because the CDP model is not nested under any of the CFA models[3], it is be possible to construct a MTMM matrix that is better fit by the CDP model than any of the CFA models. Thus, fit in this narrowly defined sense can never be used to demonstrate the absolute superiority of either approach. Also, the typically better fit of the CFA models is at the expense of estimating considerably more parameters. Whereas the TLI and Ck penalize for a lack of model parsimony, a sufficiently extreme penalty for lack of parsimony would lead to favoring the CDP model over the CFA models even for the data considered here. In summary, a limited amount of research suggests that CFA models are typically able to fit real data better than CDP models, but only at the expense of considerable parsimony.

It is also useful to compare the interpretations of the CFA models (in Appendix 2) and the CDP models (in Appendix 4) more closely. In terms of superficial support for convergent and discriminant validity, the two approaches resulted in comparable results for all five data sets considered here. Support for convergent and discriminant validity were strong for the Byrne, the YIT, and -- to a lesser extent -- the Kelly and Fiske data. Both approaches indicated good support for convergent validity but no support for discriminant validity with the Ireland data. For the Lawler data both approaches offered mixed support for convergent validity, although support for discriminant validity appeared to be stronger for the CFA approach than the CDP approach. Even this apparent difference with the Lawler data is easily explained. Estimated trait correlations for the two approaches are very similar and consistently less than 1.0. According to criteria for discriminant validity in the CFA model these results constitute support for discriminant validity whereas the CDP approach -- based on the original Campbell-Fiske guidelines -- further requires that Pt correlations are larger than Pm correlations.

While admiring the parsimony of the CDP model, it must also be recognized that this parsimony undermines much of the heuristic value of the MTMM paradigm as a formative tool. To illustrate this concern we note that there are specific features evident in the MTMM matrices that are reflected in the CFA solutions but not the CDP solutions.

1. In the Byrne data, correlations among traits are systematically lower for M1 and systematically higher for M3. This pattern is clearly evident in the sizes of correlated uniquenesses associated with each method in the CFA-CTCU model (Appendix 2) but apparently not in the CDP model (Appendix 4). Also, convergent validities in the Byrne data are consistently larger for T3. This is reflected in the higher trait factor loadings associated with T3 in the CFA-CTCU model (Appendix 2) but not in the CDP model (Appendix 4).

2. In the Lawler data, convergent validities associated with M3 (self-ratings) are low for all traits, but clearly larger for T3 (.30 and .30) than for T1 (.01 and .01) and T2 (.13 and .09). In the CFA-CTCU model this is evident in the statistically significant trait factor loading for T3M3 (.349) compared to the nonsignificant trait factor loadings for T1M3 (.095) and T2M3 (.126), but not in the CDP model.

3. For the YIT data, convergent validities associated with T4 are consistently smaller. This is reflected in the trait factor loadings in the CFA models (Appendix 2) but not the CDP model. Convergent validities associated with M2 are higher than those associated with M3 which are higher than those associated with M1. This observation is readily apparent in the Pm correlations for the CDP model (Appendix 4), but are also evident -- perhaps less obviously -- by noting differences in trait factor loadings associated with traits measured at M1, M2 and M3 (Appendix 2).

4. For the Kelly and Fiske data, convergent validities are consistently largest for T1 and lowest for T4. These patterns are evident in the CFA results (Appendix 2; except, perhaps, for the anomalous trait factor loading for T4M1) but not in the CDP solution (Appendix 4).

These more detailed comparisons of CFA and CDP solutions often revealed potentially important nuances in the data that were captured by the CFA approach but not the CDP approach. In order to illustrate this condition more clearly, we constructed an artificial MTMM matrix from a CFA-CTUM model (see Appendix 5) in which there were small method effects, small to moderate trait correlations, substantial trait variance for T1 and T2, and only weak trait variance for T3. Consistent with this design of the data, convergent validities were large for T1 and T2 (.56 to .72) but small for T3 (.09 to .15). This data should be troublesome for the CDP model that requires all convergent validities associated with a given method to be the same. Based on a hypothetical N=500, the CDP model provided an excellent fit to this artificial data (X^2 (21) = 11.15). Parameter estimates for the CDP model reflected trait correlations with a reasonable accuracy but not the large differences in convergent validities for the three traits. Furthermore, the Pm correlations -- the convergent validities -- which were all greater than .9 appear to be grossly inflated in relation to the observed convergent validities and the population model used to generate the data -- particularly given that T3 was so weak. This apparent misrepresentation of the data is particularly troublesome given the extremely good fit of the CDP model. This example, even more than the results of the 5 real data sets, demonstrates that it is important to critically evaluate parameter estimates based on different latent trait models in relation to each other and in relation to the original MTMM matrix.

In summary, this investigation has an important message for applied researchers who wish to use the MTMM paradigm. MTMM data has an inherently complicated structure that will not be fully described in all cases by any of the models or approaches considered here. There is, apparently, no "right" way to analyze MTMM data that works in all situations. Instead, we recommend that researchers consider several alternative approaches to evaluating MTMM data -- an initial inspection of the MTMM matrix using the Campbell-Fiske guidelines followed by fitting at least the subset of CFA models in Figure 1 and the CDP model. The Campbell-Fiske guidelines should be used primarily for formative purposes, the CDP seems most appropriate primarily as a summative tool, and the CFA models apparently serve both summative and formative purposes. It is, however, important that researchers understand the strengths and weaknesses of the different approaches. Despite the inherent complexity of MTMM data, we feel confident that the combination of common sense, a stronger theoretical emphasis to the design of MTMM studies, a stronger emphasis on the quality of measurement at the level of trait-method units, an appropriate arsenal of analytical tools such as recommended here, and a growing understanding of these analytic tools will allow researchers to use effectively the MTMM paradigm.

FOOTNOTES

1 -- Actually, the determinant of a variance/covariance matrix is the product of the variance components only if the covariance terms are zero so that the matrix is diagonal. In the block diagonal CCA model covariance terms involving the general factor and those relating trait contrast factors to method contrast factors are zero, but covariances among the trait contrast factors and among the method contrast factors are freely estimated. Using the Kelly and Fiske data, Wothke (1987) demonstrated that the orthogonalization of the submatrices involving trait contrast factors and method contrast factors could be accomplished by an eigenvalue decomposition like that typically conducted in principal components analysis.

2 -- It should be noted that comparisons with the Bagozzi and Yi (1990) results should be qualified in that: (a) they reported results for only the version of the CDP model in which the error structure was required to have a direct product structure -- a model that is more restrictive than the CDP model applied here and apparently inappropriate in some situations; (b) they did not consider the CFA-CTCU model emphasized here and relied primarily on the CFA-CTCM model; (c) consistent with results presented here and elsewhere, at least some of their CFA-CTCM solutions were technically improper (Bagozzi & Yi, 1990, p. 553).

3 -- Our emphasis has been on the differences between the CDP and CFA models. In general the two models are not equivalent, but it is possible for the two models to provide equivalent solutions in special circumstances. To illustrate this point, we generated a MTMM matrix that was the Kronecker product of a 3x3 Pm matrix in which all off-diagonal values were .8 and a 3x3 Pt in which all off-diagonals were .3. The CDP model, of course, provided a perfect fit for this simulated data and captured the original Pm and Pt correlations. The CFA-CTCU model, however, also fit the data perfectly as did the CFA-CTUM and CCA models (see Browne, 1984, 1989 for a mathematical derivation of the conditions under which the these models result in equivalent solutions). In the CFA-CTCU model all the squared trait-factor loadings were .8 (the off-diagonals in Pm that reflect convergent validity), whereas all the trait correlations and correlations among uniquenesses were .3 (the off-diagonals in the Pt matrix that reflect both trait effects and method effects). For other simulated data sets constructed from Pm and Pt matrices that did not have equal off-diagonal values, the CFA and CCA models were not able to perfectly fit the data demonstrating that there will be circumstances in which the CDP model is able to fit the data better than the other models.

REFERENCES

Althauser, R. P. & Heberlein, T. A. (1970). Validity and the multitrait- multimethod matrix. In E. F. Borgatta & W. Bohrnstedt (Eds.), Sociological methodology 1970 (pp. 151-169). San Francisco: Jossey-Bass.

Alwin, D. F. (1974). Approaches to the interpretation of relationships and the multitrait-multimethod matrix. In H. L. Costner (Ed.), Sociological methodology 1973-4 (pp. 79-105). San Francisco: Jossey-Bass.

Bachman, J. G. (1975). Youth in transition. Data File Documentation. Vol. 2. Ann Arbor, MI: Institute for Social Research.

Bagozzi, R. P. & Yi, L. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. Journal of Applied Psychology, 75, 547-560.

Bentler, P. M. (1990). Comparative fit indices in structural models. Psychological Bulletin, 107, .

Bock, R. D. (1960). Components of variance analysis as structural and discriminal analysis for psychological tests. British Journal of Statistical Psychology, 13, 151-163.

Bock, R. D., & Bargmann, R. E. (1966). Analysis of covariance structures. Psychometrika, 31, 507-534.

Browne, M. W. (1980). Factor analysis of multiple batteries by maximum likelihood. British Journal of Mathematical and Statistical Psychology, 33, 184-199.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. British Journal of Mathematical and Statistical Psychology, 37, 1-21.

Browne, M. W., & Cudeck, R. (1989). Single-sample cross-validation indices for covariance structures. Multivariate Behavioral Research, 24, 445-455.

Byrne, B. M., & Shavelson, R.J. (1986). On the structure of adolescent self-concept. Journal of Educational Psychology, 78, 474-481.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Campbell, D. T. & O'Connell, E. J. (1967). Method factors in multitrait- multimethod matrices: Multiplicative rather than additive? Multivariate Behavioral Research, 2, 409-426.

Campbell, D. T. & O'Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg & L. H. Kidder (Eds.), New directions for methodology of social and behavioral science: Forms of validity in research (pp. 93-111). San Francisco: Jossey-Bass.

Cudeck, R.(1988). Multiplicative models and MTMM matrices. Journal of Educational Statistics, 13, 131-147.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. Psychological Bulletin, 105, 317-327.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512-519.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. Multivariate Behavioral Research, 18, 147-167.

Fiske, D. W. (1982) Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. H. Kidder (Eds.), New directions for methodology of social and behavioral science: Forms of validity in research (pp. 93-111). San Francisco: Jossey-Bass.

Joreskog, K. G. (1971). Statistical analyses of sets of congeneric tests. Psychometrika, 36, 109-134.

Joreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2, pp. 1-56). San Francisco: W. H. Freeman.

Joreskog, K. G. (1981). Analysis of covariance structures. Scandinavian Journal of Statistics, 8, 65-92.

Joreskog, K. G., & Sorbom, D. (1989). LISREL 7: A guide to the program and applications. Chicago: SPSS, Inc.

Kelly, E. L., & Fiske, D. W. (1951). The prediction of performance in clinical psychology. Ann Arbor: University of Michigan Press.

Kavanagh, M. J., Mackinney, A. C. & Wolins, L. (1981). Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 75, 34-49.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. Journal of Experimental Social Psychology, 12, 247-252.

Kenny, D. A. (1979). Correlation and causality. New York: Wiley.

Kenny, D. A., & Kashy, D. A. (in press). The analysis of the multitrait-multimethod matrix by confirmatory factor analysis. Psychological Bulletin.

Lawler, E. E. (1969). The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 51, 369-381.

Marsh, H. W. (1982). Validity of student's evaluations of college teaching: A multitrait-multimethod analysis. Journal of Educational Psychology, 74, 264-279.

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait- multimethod data: Many problems and a few solutions. Applied Psychological Measurement, 13, 335-361.

Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), Educational research methodology, measurement and evaluation: An international handbook. Oxford, Pergamon Press.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analysis of multitrait-multimethod data: A comparison of alternative models. Applied Psychological Measurement, 15, 47-70.

Marsh, H. W., Balla, J. and McDonald (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103, 391-410.

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self-concept: Its hierarchical structure and its relation to academic achievement. Journal of Educational Psychology, 80, 366-380.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H.W. & Ireland, R. (1984). Multidimensional evaluations of writing effectiveness. (ERIC Document Reproduction Service No. ED 242 785).

Marsh, H.W. & Ireland, R. (1988). The assessment of writing effectiveness: A multidimensional perspective. Australian Journal of Psychology, 39, 353-367.

McDonald, R. P, & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. Psychological Bulletin, 107, 247-255.

Millsap, R. E. (1990). A cautionary note on the detection of method variance in multitrait-multimethod data. Journal of Applied Psychology, 75, 350-353.

Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. Psychometrika, 48, 73-83.

Schmitt, N., & Stults, D. M. (1986). Methodological review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10, 1-22.

SPSS (1989). SPSSx User's Guide (3rd E.). Chicago: SPSS.

Stanley, J. (1961). Analysis of unreplicated three-way classification with application to rater bias and trait independence. Psychometrika, 26, 205-219.

Sullivan, J. L., & Feldman, S. (1979). Multiple indicators: An introduction. Beverly Hills, Calif.: Sage Publications.

Swain, (1975). A. J. (1975). Analysis of parametric structures for variance matrices. Unpublished doctoral dissertations, University of Adelaide.

Tucker, L. R. & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. Psychological Bulletin, 74, 193-212.

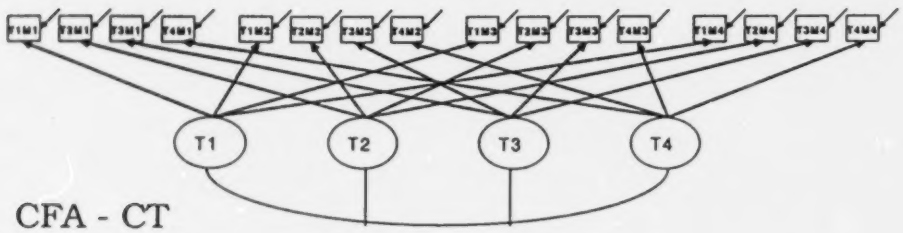
Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. Applied Psychological Measurement, 9, 1-26.

Wothke, W. (1984). The estimation of trait and method components in multitrait-multimethod measurement. (An unpublished Dissertation, University of Chicago).

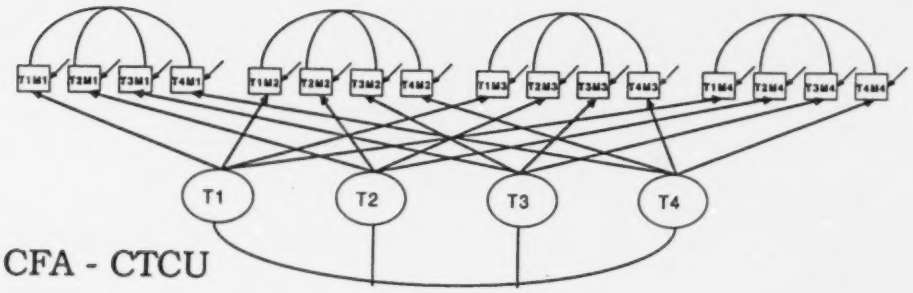
Wothke, W. (1987). The estimation of trait and method components in multitrait-multimethod measurement. ERIC

Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. Psychometrika, 55, 255-262.

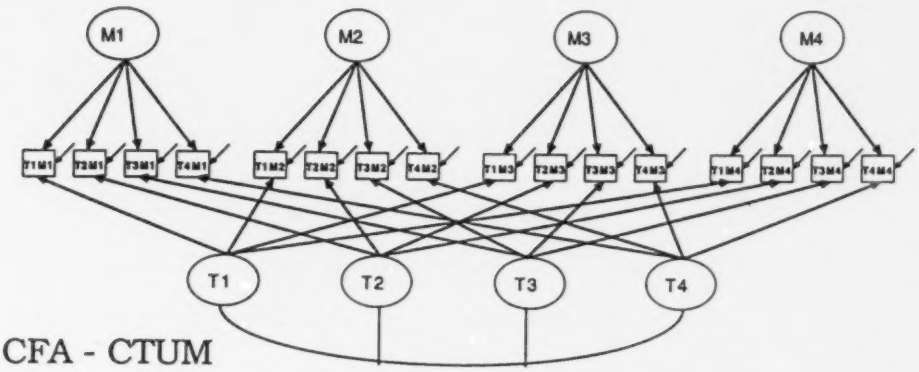
Figure 1. Four Confirmatory Factor Analysis (CFA) Models For a 4 Trait (T) x 4 Method (M) Design. Each of the 16 measured variables (T1M1, T2M1, ..., T4M4) is represented by a single measured variable (the boxes) and latent trait factors (T1-T4) and method factors (M1-M4) are represented as ovals.



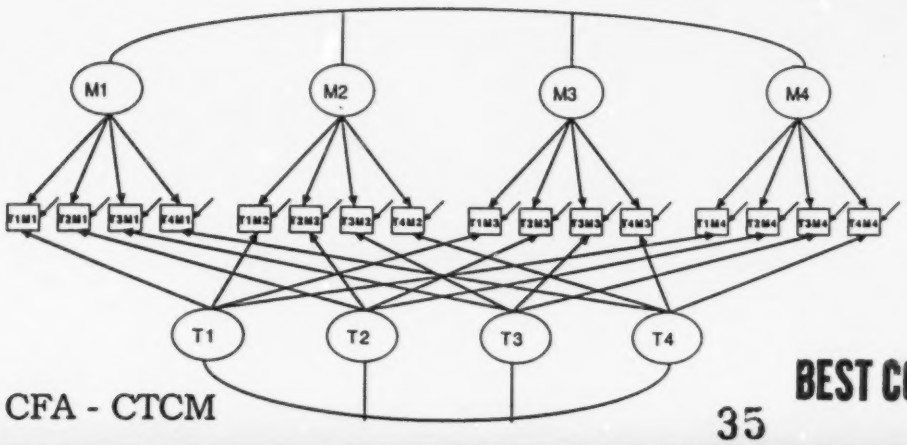
CFA - CT



CFA - CTCU



CFA - CTUM



CFA - CTCM

BEST COPY AVAILABLE

Table 1

Summary of two manifest variable approaches: The Campbell-Fiske guidelines and variance components from the ANOVA model

			Traditional Campbell-Fiske Guidelines													ANOVA Variance Components			
			Criterion 1				Criterion 2				Criterion 3				Criterion 4				
	Design		Convergent rs				HTHM rs				HTMM rs				Pattern rs				
Study	T	M	Mn	Min	Max	Crit 1	Mn	Min	Max	Crit 2	Mn	Min	Max	Crit 3	Mn	Min	Max	Conv	Disc	Meth	Error
Byrne	3	3	.70	.54	.87	9/9	.29	.01	.51	36/36	.35	.00	.58	33/36	.66	.66	.67	.45	.41	.06	.24
Lawler	3	3	.28	.01	.65	5/9	.16	.01	.42	28/36	.45	.14	.56	10/36	.65	.63	.66	.29	.13	.29	.43
Kelly	5	3	.36	.14	.71	14/15	.13	-.11	.41	111/120	.16	-.19	.46	97/120	.72	.50	.83	.19	.23	.03	.61
YIT	5	3	.52	.36	.63	15/15	.13	.00	.30	120/120	.17	.01	.40	119/120	.69	.41	.88	.22	.39	.04	.44
Ireland	5	3	.62	.43	.72	15/15	.59	.36	.74	88/120	.79	.49	.91	4/120	.51	.33	.65	.67	.03	.21	.18

Note. HTHM = Heterotrait-heteromethod correlations. HTMM = Heterotrait-monomethod correlations. The ANOVA variance components represent convergent validity, discriminant validity, method effects, and residual error respectively.

Table 2

Goodness of Fit of Alternative MTMM Models For Five Data Sets

	Byrne (T=3, M=3, N=817)						Lawler (T=3, M=3, N=113)						YIT (T=5, M=3, N=1200)						Kelly & Fiske (T=5, M3, N=124)						Ireland (T=5, M=3, N=139)					
Model	Proper	X^2	df	TLI	RNI	CK	Proper	X^2	df	TLI	RNI	CK	Proper	X^2	df	TLI	RNI	CK	Proper	X^2	df	TLI	RNI	CK	Proper	X^2	df	TLI	RNI	CK
Null	---	5310	36	.000	.000	6.522	---	348	36	.000	.000	3.258	---	5913	105	.000	.000	4.953	---	545	105	.000	.000	4.674	---	2680	105	.000	.000	19.524
CFA-CT	Yes	452	24	.878	.919	.605	Yes	101	24	.630	.754	1.305	Yes	510	80	.903	.926	.493	Yes	140	80	.820	.863	1.880	No	724	80	.671	.750	5.870
CFA-CTUM	No	78	15	.971	.988	.170	No	20	15	.961	.984	.765	Yes	143	65	.978	.987	.212	Yes	76	65	.959	.975	1.642	Yes	123	65	.964	.977	1.787
CFA-CTCM	No	33	12	.988	.996	.123	No	5	12	1.070	1.023	.688	Yes	123	62	.982	.989	.201	No	57	62	1.017	1.010	1.547	No	97	62	.977	.986	1.648
CFA-CTCU	Yes	78	15	.971	.988	.170	Yes	20	15	.962	.984	.765	Yes	79	50	.990	.995	.184	Yes	68	50	.916	.960	1.853	Yes	64	50	.988	.994	1.614
CCA	Yes	177	21	.949	.970	.276	Yes	29	21	.955	.974	.729	Yes	319	77	.943	.958	.338	Yes	105	77	.914	.937	1.648	No	158	77	.957	.968	1.845
CDP	No	172	21	.951	.971	.270	Yes	29	21	.957	.975	.726	Yes	149	77	.983	.988	.196	Yes	116	77	.879	.912	1.738	Yes	173	84	.949	.963	1.949
CDP-KE[a]	Yes	249	25	.939	.958	.354																								
CCA-Diag[b]																									Yes	145	77	.964	.974	1.745

Note. TLI = Tucker-Lewis Index, RNI = Relative noncentrality index, Ck = Cross-validition index. See Figure 1 for a description of the models.

[a] The Composite Direct Product Model with Kronecker Errors (CDP-KE) Model was fit to the Byrne data because the CDP model resulted in an improper solution for the Byrne data. [b] The Covariance components analysis completely diagonal model (CCA-Diag) was fit because the CCA model resulted in an improper solution for the Ireland data.

39

38

BEST COPY AVAILABLE

Appendix 1

Five MTMM correlation matrices used in this study.

Byrne data (3Tx3M)

```
t1m1  (.89)
t2m1  .384 (.79)
t3m1  .441 .002 (.92)
t1m2  .662 .368 .353  (.84)
t2m2  .438 .703 .008  .441 (.89)
t3m2  .465 .069 .871  .424 .136 (.95)
t1m3  .678 .331 .478  .550 .380 .513  (.87)
t2m3  .458 .541 .057  .381 .658 .096  .584 (.90)
t3m3  .414 .027 .825  .372 .029 .810  .582 .135 (.94)
```

Lawler data (3tx3m)

```
t1m1 1.00
t2m1  .53 1.00
t3m1  .56  .44 1.00
t1m2  .65  .38  .40 1.00
t2m2  .42  .52  .30  .56 1.00
t3m2  .40  .31  .53  .56  .40 1.00
t1m3  .01  .01  .09  .01  .17  .10 1.00
t2m3  .03  .13  .03  .04  .09  .02  .43 1.00
t3m3  .06  .01  .30  .02  .01  .30  .40  .14 1.00
```

Youth In Transition (YIT) data (5Tx3M)

```
t1m1  1.000
t2m1  .162 1.000
t3m1  .212 .085 1.000
t4m1  .256 .119 .401 1.000
t5m1  .292 .015 .054 .135 1.000
t1m2  .525 .137 .120 .216 .231  1.000
t2m2  .163 .588 .088 .144 .043  .153 1.000
t3m2  .136 .050 .488 .215 .020  .206 .058 1.000
t4m2  .186 .109 .213 .444 .098  .299 .104 .283 1.000
t5m2  .226 .022 .044 .102 .567  .346 .050 .076 .144 1.000
t1m3  .483 .123 .103 .181 .192  .633 .126 .162 .244 .278  1.000
t2m3  .141 .502 .156 .157 .000  .128 .549 .099 .096 .027  .097 1.000
t3m3  .094 .046 .416 .158 .045  .145 .006 .610 .208 .051  .195 .027 1.000
t4m3  .120 .076 .170 .365 .097  .236 .065 .229 .507 .091  .303 .080 .287 1.000
t5m3  .231 .068 .050 .151 .505  .296 .107 .043 .163 .632  .398 .071 .054 .149 1.000
```

Appendix 1 (continued)

Five MTMM correlation matrices used in this study.

KELLY and FISKE data (5Tx3M)

t1m1	1														
t2m1	.37	1													
t3m1	-.24	-.14	1												
t4m1	.25	.46	.08	1											
t5m1	.35	.19	.09	.31	1										
t1m2	.71	.35	-.18	.26	.41	1									
t2m2	.39	.53	-.15	.38	.29	.37	1								
t3m2	-.27	-.31	.43	-.06	.03	-.15	-.19	1							
t4m2	.03	-.05	.03	.20	.07	.11	.23	.19	1						
t5m2	.19	.05	.04	.29	.47	.33	.22	.19	.29	1					
t1m3	.48	.31	-.22	.19	.12	.46	.36	-.15	.12	.23	1				
t2m3	.17	.42	-.10	.10	-.03	.09	.24	-.25	-.11	-.03	.23	1			
t3m3	-.04	-.13	.22	-.13	-.05	-.04	-.11	.31	.06	.06	-.05	-.12	1		
t4m3	.13	.27	-.03	.22	-.04	.10	.15	.00	.14	-.03	.16	.26	.11	1	
t5m3	.37	.15	-.22	.09	.26	.27	.12	-.07	.05	.35	.21	.15	.17	.31	1

Ireland data (5Tx3M)

t1m1	1.000														
t2m1	.86	1.000													
t3m1	.86	.85	1.000												
t4m1	.80	.81	.89	1.000											
t5m1	.85	.84	.91	.90	1.000										
t1m2	.69	.65	.65	.63	.66	1.000									
t2m2	.68	.67	.65	.63	.66	.81	1.000								
t3m2	.71	.68	.72	.70	.74	.75	.77	1.000							
t4m2	.66	.63	.69	.69	.66	.76	.81	.83	1.000						
t5m2	.69	.68	.70	.67	.71	.84	.86	.84	.88	1.000					
t1m3	.60	.52	.56	.56	.61	.63	.54	.58	.55	.59	1.000				
t2m3	.58	.60	.55	.54	.61	.45	.57	.51	.48	.52	.63	1.000			
t3m3	.63	.62	.61	.57	.63	.55	.55	.63	.54	.56	.68	.72	1.000		
t4m3	.53	.56	.50	.53	.54	.36	.45	.48	.43	.42	.49	.74	.76	1.000	
t5m3	.60	.61	.57	.57	.60	.52	.55	.61	.55	.57	.71	.73	.88	.77	1.000

Appendix 2

Parameter Estimates From the Best Fitting Confirmatory Factor Analysis (CFA) Models

---

Byrne Data (CFA-CTCU)

	Trait Factor	unique	SMC	Unique. Correlations				
T1M1	.869*	.246*	.754	1.000				
T2M1	.775*	.394*	.604	-.119*	1.000			
T3M1	.942*	.113*	.887	-.019	-.114*	1.000		
T1M2	.731*	.463*	.536	1.000				
T2M2	.863*	.228*	.765	.130*	1.000			
T3M2	.930*	.142*	.859	.125*	.499*	1.000		
T1M3	.754*	.399*	.588	1.000				
T2M3	.755*	.450*	.558	.537*	1.000			
T3M3	.847*	.242*	.748	.423*	.214*	1.000		

Trait Correlations

T1	1.000		
T2	.604*	1.000	
T3	.596*	.042	1.000

Lawler Data: CFA-CTCU

	Trait Factor	unique	SMC	Unique. Correlations				
T1M1	.868*	.240	.759	1.000				
T2M1	.761*	.414*	.583	.251	1.000			
T3M1	.781*	.390*	.610	.341	.268	1.000		
T1M2	.730*	.454*	.540	1.000				
T2M2	.672*	.544*	.454	.428*	1.000			
T3M2	.691*	.519*	.479	.449*	.263*	1.000		
T1M3	.095	1.003*	.009	1.000				
T2M3	.126	.982*	.016	.427*	1.000			
T3M3	.349*	.819*	.123	.407*	.154	1.000		

Trait Correlations

T1	1.000		
T2	.680*	1.000	
T3	.652*	.532*	1.000

Appendix 2 (continued)
Parameter Estimates From the Best Fitting Confirmatory Factor Analysis (CFA) Models

---

YIT data: CFA-CTCU

	Trait Factor	unique	SMC	Unique.	Correlations			
T1M1	.639*	.596*	.406	1.000				
T2M1	.735*	.459*	.541	.051	1.000			
T3M1	.579*	.662*	.336	.193*	-.008	1.000		
T4M1	.572*	.668*	.329	.177*	-.014	.368*	1.000	
T5M1	.674*	.546*	.454	.210*	-.006	.040	.073	1.000
T1M2	.833*	.307*	.693	1.000				
T2M2	.797*	.364*	.636	-.002	1.000			
T3M2	.846*	.286*	.714	.089	.038	1.000		
T4M2	.772*	.403*	.597	.067	-.011	.031	1.000	
T5M2	.837*	.302*	.699	.129*	-.040	.121*	.041	1.000
T1M3	.755*	.425*	.573	1.000				
T2M3	.682*	.530*	.467	-.036	1.000			
T3M3	.724*	.481*	.521	.150*	-.080	1.000		
T4M3	.653*	.578*	.425	.215*	.004*	.197*	1.000	
T5M3	.760*	.425*	.576	.315*	.030*	.006	.049	1.000

Trait Correlations

T1	1.000				
T2	.232*	1.000			
T3	.247*	.104*	1.00 0		
T4	.413*	.194*	.420 *	1.000	
T5	.438*	.087*	.067 *	.222*	1.000

YIT: CFA-CTCM

	Trait Factor	Method Factor	Unique	SMC
T1M1	.751*	.224*	.380*	.618
T2M1	.737*	-.013	.458*	.543
T3M1	.575*	.519*	.400*	.600
T4M1	.561*	.497*	.431*	.566
T5M1	.706*	.077*	.494*	.505
T1M2	.687*	.538*	.251*	.752
T2M2	.798*	.010	.362*	.638
T3M2	.820*	.187*	.296*	.705
T4M2	.734*	.225*	.408*	.591
T5M2	.798*	.263*	.296*	.705
T1M3	.633*	.687*	.142*	.860
T2M3	.686*	-.004	.530*	.470
T3M3	.723*	.183*	.447*	.554
T4M3	.631*	.269*	.527*	.472
T5M3	.731*	.321*	.370*	.638

Trait factor Correlations

T1	1.000				
T2	.255*	1.000			
T3	.192*	.100*	1.000		
T4	.332*	.195*	.409*	1.000	
T5	.404*	.082*	.038	.175*	1.000

Method factor Correlations

M1	1.000		
M2	.178*	1.000	
M3	.116	.571*	1.000

Appendix 2 (continued)
Parameter Estimates From the Best Fitting Confirmatory Factor Analysis (CFA) Models

---

Kelly and Fiske Data: CFA-CTUM

	Trait Factor	Method Factor	Unique	SMC
T1M1	.857*	-.023	.264*	.736
T2M1	.827*	.090	.305*	.694
T3M1	.560*	.712*	.184*	.817
T4M1	.933*	.201	.087*	.913
T5M1	.695*	.214	.492*	.518
T1M2	.830*	.137	.297*	.704
T2M2	.696*	.320*	.454*	.563
T3M2	.743*	.249*	.357*	.632
T4M2	.185*	.642*	.547*	.449
T5M2	.646*	.365*	.428*	.562
T1M3	.551*	.105	.681*	.316
T2M3	.421*	.261*	.743*	.248
T3M3	.419*	.295*	.755*	.258
T4M3	.301*	.591*	.592*	.426
T5M3	.556*	.570*	.420*	.602

Trait factor Correlations

T1	1.000				
T2	.568*	1.000			
T3	-.368*	-.487*	1.000		
T4	.339*	.547*	-.120	1.000	
T5	.562*	.263*	-.007	.411*	1.000

Kelly and Fiske Data: CFA-CTCU

	Trait Factor	unique	SMC	Unique.	Correlations			
T1M1	.876*	.230*	.770	1.00				
T2M1	.827*	.318*	.682	-.161	1.00			
T3M1	.584*	.663*	.340	-.027	.153	1.000		
T4M1	.921*	.150	.849	-.148	.256	.481*	1.000	
T5M1	.690*	.541*	.468	-.041	-.014	.265	.119	1.000
T1M2	.827*	.338*	.669	1.000				
T2M2	.705*	.550*	.475	.062	1.00			
T3M2	.704*	.467*	.515	.170	.111	1.000		
T4M2	.188	.956*	.036	.141	.318*	.229*	1.000	
T5M2	.626*	.585*	.401	.221	.199	.245	.304*	1.000
T1M3	.556*	.696*	.308	1.000				
T2M3	.410*	.815*	.171	.137	1.000			
T3M3	.420*	.842*	.173	.012	-.038	1.000		
T4M3	.300*	.943*	.087	.099	.223*	.171	1.000	
T5M3	.558*	.733*	.298	.031	.169	.277*	.392*	1.000

Trait factor Correlations

T1	1.000				
T2	.591*	1.000			
T3	-.391*	-.475*	1.000		
T4	.354*	.522*	-.134	1.000	
T5	.554*	.289*	-.056	.425*	1.000

Appendix 2 (continued)
Parameter Estimates From the Best Fitting Confirmatory Factor Analysis (CFA) Models

Ireland Data: CFA-CTCU

	Trait Factor	unique.	SMC	Unique. Correlations				
T1M1	.861*	.252*	.747	1.000*				
T2M1	.835*	.290*	.706	.600*	1.000			
T3M1	.835*	.291*	.706	.588*	.596*	1.000		
T4M1	.843*	.274*	.722	.423*	.490*	.694*	1.000	
T5M1	.868*	.238*	.760	.473*	.454*	.687*	.738*	1.000
T1M2	.801*	.360*	.640	1.000				
T2M2	.799*	.372*	.632	.570*	1.000			
T3M2	.853*	.274*	.727	.320*	.411*	1.000		
T4M2	.801*	.367*	.636	.475*	.611*	.551*	1.000	
T5M2	.819*	.338*	.665	.617*	.646*	.505*	.741*	1.000
T1M3	.728*	.478*	.526	1.000				
T2M3	.690*	.517*	.480	.362*	1.000			
T3M3	.739*	.464*	.541	.360*	.502*	1.000		
T4M3	.616*	.632*	.375	.153*	.595*	.623*	1.000	
T5M3	.713*	.498*	.505	.456*	.499*	.761*	.618*	1.000

Trait Factor Correlations

T1	1.000			
T2	.948*	1.000		
T3	.960*	.948*	1.000	
T4	.928*	.932*	.963*	1.000
T5	.964*	.978*	.991*	.960* 1.000

Ireland Data: CFA-CTUM

	Trait Factor	Method Factor	Unique	SMC
T1M1	.864*	.329*	.138*	.861
T2M1	.836*	.363*	.160*	.838
T3M1	.837*	.468*	.072*	.928
T4M1	.837*	.434*	.098*	.901
T5M1	.860*	.407*	.086*	.914
T1M2	.801*	.403*	.201*	.800
T2M2	.807*	.454*	.153*	.849
T3M2	.862*	.288*	.179*	.822
T4M2	.819*	.470*	.124*	.878
T5M2	.828*	.508*	.069*	.932
T1M3	.713*	.297*	.400*	.599
T2M3	.683*	.449*	.322*	.675
T3M3	.735*	.588*	.122*	.879
T4M3	.608*	.586*	.295*	.707
T5M3	.714*	.616*	.114*	.887

Trait factor correlations

T1	1.000				
T2	.965*	1.000			
T3	.958*	.942*	1.000*		
T4	.912*	.929*	.967*	1.000	
T5	.964*	.964*	.988*	.969*	1.000

Note. CFA results are summarized for the best models for each data set. Each measured variable is a trait-method unit. T1M1, for example, is trait 1 measured by method 1. The squared multiple correlations (SMC) are an estimate of the communality for each measured variable.

* $p < .05$

Appendix 3

Parameter Estimates For the Covariance Components Analysis (CCA) Models

---

Byrne Data

Variance/covariances

```
 GEN 1.000
 TC1   0  .074*
 TC2   0  .024  .589*
 MC1   0   0     0    .022*
 MC1   0   0     0    .017* .069*
```

Squared multiple correlations

```
 T1M1 T2M1 T3M1 T1M2 T2M2 T3M3 T1M3 T2M2 T3M3
 .730 .601 .881 .565 .831 .905 .849 .722 .826
```

Lawler data

Variance/covariances

```
 GEN 1.000
 TC 1  0   .102*
 TC2   0   .016  .246*
 MC1   0    0     0   .282*
 MC1   0    0     0   .271* .579*
```

Squared multiple correlations

```
 T1M1 T2M1 T3M1 T1M2 T2M2 T3M3 T1M3 T2M2 T3M3
 .810 .626 .672 .821 .618 .664 .563 .270 .399
```

YIT data

Variance/covariances

```
 GEN   1.000
 TC1     0    .205*
 TC2     0   -.070*   .661*
 TC3     0   -.062*  -.027  .380*
 TC4     0   -.036*   .037  .147*  .365*
 MC1     0     0       0     0      0    .105*
 MC2     0     0       0     0      0    .020* .051*
```

Squared multiple correlations

```
 T1M1 T2M1 T3M1 T4M1 T5M1 T1M1 T2M1 T3M1 T4M1 T5M1 T1M1 T2M1 T3M1 T4M1 T5M1
 .535 .516 .438 .455 .507 .704 .653 .712 .586 .709 .710 .446 .571 .504 .640
```

Kelly and Fiske data (T3 negatively oriented)

Variance/covariances

```
 GEN 1.0
 TC1  .0   .670*
 TC2  .0   .387*   .793*
 TC3  .0  -.639*  -.821* 1.2844*
 TC4  .0  -.087    .107   .020  .169*
 MC1  .0    0       0      0     0   .126*
 MC2  .0    0       0      0     0   .100* .360*
```

Squared multiple correlations

```
 T1M1 T2M1 T3M1 T4M1 T5M1 T1M1 T2M1 T3M1 T4M1 T5M1 T1M1 T2M1 T3M1 T4M1 T5M1
 .735 .667 .370 .589 .467 .738 .535 .591 .166 .566 .263 .276 .201 .449 .648
```

Appendix 3 (continued)
Parameter Estimates For the Covariance Components Analysis (CCA) Models

---

Kelly and Fiske data (T3 reflected so as to be positively oriented)

Variance/covariances

GEN	1.0					
TC1	.0	.142*				
TC2	.0	-.008	.101*			
TC3	.0	-.045	.006	.790*		
TC4	.0	-.067*	.086*	.094	.204*	
MC1	.0	0	0	0	0	.022
MC2	.0	0	0	0	0	.046	.115*

Squared multiple correlations

T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1
.704	.692	.278	.519	.449	.727	.461	.678	.069	.473	.398	.311	.107	.228	.310

Ireland data: CCA (fully diagonal since block diagonal ill defined)

Variance/covariances

GEN	1.000					
TC1	0	.013*				
TC2	0	.000	.013*			
TC3	0	.000	0	.002		
TC4	0	.000	0	.000	.003	
MC1	0	.000	0	.000	0	.089*
MC2	0	.000	0	.000	0	.000	.139*

Squared multiple correlations

T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1
.855	.841	.919	.873	.917	.799	.853	.792	.854	.927	.594	.689	.877	.686	.885

Note. CCA results are summarized as the variance/covariance matrix for the general (GEN) contrast factor, the trait contrast (TC1, TC2,..) factors, and the method contrast (MC1, MC2, ...) factors. The squared multiple correlations (SMC) are an estimate of the communality for each measured variable.

* $p < .05$

Appendix 4

Parameter Estimates For the Composite Direct Product (CDP) Models

---

Byrne Data: CDP Model

Trait Corr			Method Corrs		
T1	1		M1	1	
T2	.683*	1	M2	.955*	1
T3	.600*	.181* 1	M3	.851*	.814* 1

Squared multiple correlations

T1M1	T2M1	T3M1	T1M2	T2M2	T3M3	T1M3	T2M2	T3M3
.756	.610	.889	.586	.867	.933	.856	.851	(1.0)

Byrne Data: CDP-CE Model (because CDP solution was improper)

Trait Corr			Method Corrs		
T1	1		M1	1	
T2	.694*	1	M2	.879*	1
T3	.604*	.161* 1	M3	.853*	.840* 1

Squared multiple correlations

T1M1	T2M1	T3M1	T1M2	T2M2	T3M3	T1M3	T2M2	T3M3
.665	.700	.896	.673	.707	.900	.850	.867	.961

Lawler data: CDP Model

Trait correlations				Method correlations			
T1	1.000			M1	1.000		
T2	.687*	1.000		M2	.717*	1.000	
T3	.665*	.520*	1.000	M3	.207*	.190*	1.000

Squared multiple correlations

T1M1	T2M1	T3M1	T1M2	T2M2	T3M3	T1M3	T2M2	T3M3
.878	.710	.782	.902	.724	.748	.802	.454	.541

YIT data: CDP Model

t Correlations					Method Correlations			
1.000					M1	1.000		
.123*	1.000				M2	.685*	1	
.272*	.026	1.000			M3	.579*	.782*	1
.396*	.086*	.471*	1	.000				
.425*	.036	.072*	.179*	1 .000				

Squared multiple correlations

T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1
.728	.999	.779	.709	.830	.757	.699	.763	.631	.768	.867	.687	.792	.644	.865

Kelly and Fiske Data: CDP Model

Trait correlations					Method correlations			
T1	1.000				M1	1.000		
T2	.442*	1.000			M2	.796*	1.000	
T3	-.171	-.212*	1.000		M3	.579*	.529*	1.000
T4	.295*	.636*	.253*	1.000				
T5	.404*	.299*	.309*	.589*	1.000			

Squared multiple correlations

T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1
.830	.724	.398	.667	.535	.830	.604	.787	.213	.710	.511	.426	.465	.515	.747

Ireland data: CDP model

Trait correlations					Method correlations			
T1	1.00				M1	1.000		
T2	.952*	1.000			M2	.767*	1.000	
T3	.935*	.932*	1.000		M3	.673*	.652*	1.00
T4	.877*	.917*	.964*	1.000				
T5	.945*	.956*	.989*	.970*	1.000			

Squared multiple correlations

T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1	T1M1	T2M1	T3M1	T4M1	T5M1
.881	.851	.920	.913	.902	.814	.860	.798	.892	.914	.603	.700	.885	.719	.878

Note. Each measured variable is a trait-method unit. T1M1, for example, is trait 1 measured by method 1. The squared multiple correlations (SMC) are an estimate of the communality for each measured variable.

* p < .05

BEST COPY AVAILABLE

Appendix 5

Simulated Data Used to Compare CFA and CDP Models

CFA-CTUM Model Used

To Generate the Data

	TFL	MFL	TD
T1M1	.9	.1	.18
T2M1	.8	.1	.35
T3M1	.3	.3	.82
T1M2	.8	.2	.32
T2M2	.8	.2	.32
T3M2	.3	.3	.82
T1M3	.7	.1	.50
T2M3	.7	.2	.47
T3M3	.5	.3	.66

Trait factor Correlations

T1	1.0		
T2	.6	1.0	
T3	.1	.1	1.0

Artificial MTMM Matrix

t1m1	1.000								
t2m1	.442	1.000							
t3m1	.057	.054	1.000						
t1m2	.720	.384	.024	1.000					
t2m2	.432	.640	.024	.424	1.000				
t3m2	.027	.024	.090	.084	.084	1.000			
t1m3	.630	.336	.021	.560	.336	.021	1.000		
t2m3	.378	.560	.021	.336	.560	.021	.314	1.000	
t3m3	.045	.040	.150	.040	.040	.150	.065	.095	1.000

Artificial data: CDP Model

Trait Correlations			Method Correlations		
T1	1.000		M1	1.000	
T2	.597	1.000	M2	.934	1.000
T3	.160	.179	1.000	M3	.943	.904	1.000

Squared multiple correlations

T1M1	T2M1	T3M1	T1M2	T2M2	T3M3	T1M3	T2M2	T3M3
.832	.656	.098	.713	.714	.117	.536	.538	.246

Note. The CFA-CTUM solution was used to generate the MTMM matrix that was then evaluated with the CDP model. Even though the true trait variances for T1 and T2 differed substantially from T3, apparently violating an assumption of the CDP model, the CDP model fit the data very well (X^2 (21) = 11.15 for a hypothetical N=500).